

# THESE DE DOCTORAT DE L'UNIVERSITE PARIS-SUD

Ecole Doctorale : *Gènes, Génomes Cellules - ED426*  
Laboratoire *Biologie Moléculaire du Gène chez les Extrêmophiles,*  
*Institut de Génétique et de Microbiologie*

DISCIPLINE Biologie

présentée par  
**Ludovic PERROCHIA**



## **Caractérisation Biochimique des Machineries de Biosynthèse de t<sup>6</sup>A, un Nucléoside Modifié Universel**



**soutenue le 25 Juin 2013**



### Composition du jury

<b>Patrick FORTERRE</b>	Professeur, Université Paris Sud	<b>Directeur de thèse</b>
<b>Valérie DE CRECY-LAGARD</b>	Associate Professor, University of Florida	<b>Rapporteur</b>
<b>Hannu MYLLYKALLIO</b>	Directeur de recherche, CNRS	<b>Rapporteur</b>
<b>Iouri MOTORINE</b>	Professeur, Université de Lorraine	<b>Examineur</b>
<b>Jean Pierre ROUSSET</b>	Professeur, Université Paris Sud	<b>Examineur</b>
<b>Tamara BASTA</b>	Maître de Conférences, Université Paris Sud	<b>Invité</b>
<b>Arnaud HECKER</b>	Maître de Conférences, Université de Lorraine	<b>Invité</b>



# Remerciements

Je tiens à remercier chaleureusement les membres de mon jury de thèse qui ont bien voulu juger et évaluer mon travail: mes rapporteurs, **Hannu Myllykallio**, qui m'a également suivi au sein de mon comité de thèse, et **Valérie de Crécy-Lagard** qui ont eu le courage de lire ces 300 et quelques pages, ainsi que **Iouri Motorine** et **Jean-Pierre Rousset**.

Merci avant tout à **Patrick**, le *Gandalf* des Archées à la fois plein de sagesse et de folie. Merci de m'avoir accueilli si chaleureusement au sein de ta *Communauté*, d'avoir été mon mentor en acceptant de diriger ma thèse, d'avoir partagé ta passion, tes points de vues, tes idées et tes conseils. Ta vision du vivant et de la Science, parfois (souvent !) surprenante, a profondément modifié mes propres conceptions, et j'espère pouvoir transmettre à mon tour cette vision aussi bien que tu me l'as transmise. Merci également de m'avoir accordé pleinement ta confiance, non seulement sur le plan scientifique, mais aussi en me proposant d'assurer à ta place les différents enseignements sur les Archées, et de m'avoir permis de traverser la *Grande Mer de l'Ouest* pour faire ce pèlerinage inoubliable sur la Terre sacrée des *Elks* et des hyperthermophiles (*i.e.* Yellowstone) ; sans parler des autres congrès auxquels j'ai eu la chance de participer.

A mes deux *Maîtres Jedi* de la Recherche, qui ont su me guider pendant ces 4 ans à travers les chemins tortueux de la biochimie, j'espère avoir été un bon *Padawan*. A **Arnaud**, qui m'a formé pendant un an en me transmettant la passion de la biochimie, m'a livré tous les secrets de la purification de protéines, m'a appris à maîtriser la FPLC, et m'a véritablement donné envie de me lancer dans cette thèse, et qui a continué à me suivre après son départ pour la « province » ;). A **Tamara**, qui a ensuite pris le relais et qui m'a permis de passer de l'étudiant au véritable scientifique. Je n'aurais pas pu espérer une meilleure chef ! Merci pour toutes ces discussions, ces conseils, les échanges d'idées que nous avons eus, pour la confiance que tu m'as donnée et ta gentillesse au quotidien, merci d'avoir patiemment corrigé mes présentations et mon manuscrit, d'avoir tenté de m'apprendre à être organisé, et stresser à ma place quand je faisais les choses à la dernière minute (promis j'essayerai de ne plus le faire, même si c'est totalement irréaliste...). Je peux dire sincèrement que j'ai vraiment eu de la chance, et que tu as largement contribué à ce que cette thèse soit pour moi une période géniale ! Désolé donc à mes deux Maîtres Jedi si je passe *du côté obscur de la Force* (*i.e.* l'enseignement !), mais je garderai grâce à vous un super souvenir de la recherche.

**Un grand merci également à toutes les personnes avec qui j'ai travaillé pendant ces années et qui ont participé directement ou indirectement à cette thèse. Merci donc...**

... à toutes les victimes de la malédiction de KEOPS: **Estelle**, qui a su dompter la première le scintillateur et a réussi à faire marcher une réaction qui semblait si improbable. Merci également pour ce *road trip* américain, pour son *self control* face aux loups et aux elks et pour les super photos (même si les miennes sont quand même mieux !!!). **Dorian**, qui a torturé ce pauvre KEOPS en le mutant et en le disséquant, nous permettant d'en révéler tous les secrets.

Nos collaborateurs : l'équipe des cristallos, **Joseph**, **Bruno**, **Wenhua**, **Herman**, et l'équipe de la Levure, **Marie-Claire** et **Domenico**, pour leur discussion, les échanges d'idées, d'ARN et de protéines, qui me permettent de dire « oui, j'ai aussi travaillé chez les Eucaryotes ! »

... à mes deux inséparables co-bureaux **Jack Daniel's**, pour leur bonne humeur au quotidien, et pour avoir supporté mon étalement progressif et entropique sur la quasi-totalité des paillasse et du bureau, les traces de dents sur tous les marqueurs du labo, et mes emprunts répétés du thermomixeur. **Jacques**, pour ses blagues plus ou moins réussies, mais aussi ses discussions et ses conseils. En revanche, un grand « anti-merci » pour son petit air de sifflement qui je pense va rester dans ma tête pendant encore un certain temps ! **Danièle**, avec qui j'ai partagé les galères de purif, et qui a partagé avec moi la lourde tâche de prendre soin de nos colonnes, a pris le relais de s'en occuper quand elles étaient au plus mal, et qui m'a montré que les topoisomérases, ça pouvait être en fin de compte intéressant !

... à la seule et unique **Evelyne**, qu'il faudrait inventer si elle n'existait pas. Pour sa joie de vivre communicative, sa générosité, sa gentillesse (et aussi ses coups de gueule ! ;), ses anecdotes, et pour s'être occupé de la paperasse que personne n'a jamais envie de faire.

... aux **Topoisoméristes** **Marc**, **Florence** et **Hélène**, pour leurs conseils lors des labmeeting et des répétitions des différentes présentations orales que j'ai dû faire.

...à **Myriam**, la petite main invisible de l'équipe, sans qui aucun travail ne serait possible.

...aux **Drôles de Dames de la Radioactivité**, qui m'ont permis pendant ces trois ans de travailler avec tout un tas de substances radioactives en toute sécurité. **Claire**, la « *Danièle et Béatrice* » de la radioc, pour ses innombrables frottis et sa gentillesse (même quand elle me grondait !), qui m'a présenté à mon plus terrible ennemi : le scintillateur. **Martine**, ma dealeuse de radioc, pour m'avoir fourni ma dose presque mensuelle d'ATP mais aussi ma dose quotidienne de café (et de sucre !), pour sa bonne humeur à midi. **Marie Claude**, la respo chef radioc, pour ses idées, ses discussions, et pour m'avoir fait participer à la Fête de la science.

...aux **spécialistes des ARNt**: **Henri Grosjean**, le *Maître Yoda* des modifications d'ARNt et de  $t^6A$ , pour ses conseils et ses suggestions lors de l'écriture de notre premier article, et pour toute la biblio et ses connaissances sur  $t^6A$  qu'il a partagé avec nous, et **Florence**, la reine des yaourts, qui a accepté de faire partie de mon comité de thèse, qui m'a conseillé, et dont la thèse m'a servi d'inspiration.

...à mes co-thésardes **Archaealogistes**. **Mohea** (ou Moya/Moa/Moeua), la MacGiver des Sulfo, qui m'a accepté et accueilli (même en sachant mes propensions naturelles au désordre) dans son antre et avec qui j'ai partagé les joies de la rédaction, pour nos discussions, nos délires, et bien sûr pour ce méga voyage à Grenoble ;) **Marie**, la pro des vésicules et future collègue, pour s'être occupé de mes petites plantes pendant mes absences, pour les soirées chez elle, et surtout pour sa redoutable clochette anti-ours qui nous a permis de rentrer sains et saufs de Yellowstone.

... et à toutes les autres personnes de l'IGM avec qui j'ai discuté et échangé durant mon passage ici.

Et comme une thèse, ce n'est pas uniquement une vie enfermée dans un laboratoire, je tiens à remercier tous mes amis et proches qui m'ont permis de tenir psychologiquement durant ces longues années d'exil au pays du froid et de la grisaille. Merci donc...

...aux **Eucaryologistes Lyonnaises**, qui pensent qu'il faut avoir un noyau pour mériter l'appellation « cellule »! (et à qui du coup ont intérêt à lire mon intro ;) **Marina**, qui m'a fait découvrir le Mercalm pendant notre mémorable expédition en Corse, pour les petites photos ponctuelles « remonte-moral » envoyées sur mon téléphone. **MMM's**, pour avoir montré qu'on peut être maman mais rester une pote comme si de rien n'était, et qui est l'exemple même de la bonne humeur constante! **Dr Marion**, pour ses cocktails improbables et les looongues heures de marche à travers Manhattan. **Charlène**, qui m'a permis récemment de découvrir une facette insoupçonnée de moi-même!



**...aux (exilés) Niçois.** La **Cynth'**, ma meilleure amie, qui a rendu *trop canon* chacun de mes retours à Nice, qui m'a accompagné à la plage pendant ces 4 ans et qui m'a permis ne pas finir blanc comme un ... culot de protéines. **Ben**, à qui je fais une spéciale dédicace figure 3, pour ce fantastique voyage en bateau marseillais et le Kirby qui me tient compagnie à Orsay. **Yannick**, pour cette expédition au pays des volcans, et qui m'a finalement rejoint dans cette somptueuse et ensoleillée région parisienne, me permettant de profiter des Happy hours parisiens. **Lola**, la seule future vraie docteur de la bande, qui m'a donné l'occasion de descendre dans le sud, mais pas pour Nice ! **Matthieu** le troubadour et **Amélie** la lettrée, qui sont venus me rendre visite et ont fait connaissance avec la grande cité d'Orsay et ses habitants étranges, et **Coconut**, qui était là à chacune des meilleures soirées niçoises.

**...aux autres amis :** les bulbophiles **Marie** et **Elsa** pour ce super trip à Amsterdam, **Souhir** qui a partagé avec moi ma première année au labo, **Pauline** et **Aurore**, les compères de Yellowstone et voleuses d'obsidienne, sans oublier **Béa**, **Olesya**, **Olivier**, **Anthony**, **Pierre**, **Jérôme** et la mini **Gaga**.

**...à ma famille :** qui a fait l'effort d'essayer de comprendre ce que je trafiquais au laboratoire pendant ces 4 ans : **Mum**, sans qui cette thèse serait truffée d'horribles fêtes d'ortographe, **Hib**, **Dolly**, **Momo**, le petit **Jibou**, et une petite pensée pour **Mutti**.

Enfin, au **Dr Djo**, qui a été là à chaque moment pendant ces 4 ans, et qui est devenue malgré elle une spécialiste de KEOPS. Merci pour ta patience, pour avoir écouté chacune de mes idées et y avoir réfléchi avec moi, même si c'était à dix mille années lumière de ton domaine, d'avoir partagé à la fois mes joies, mes espoirs et mes déceptions, de m'avoir conseillé quand j'avais des difficultés alors que tu avais tes propres problèmes, d'avoir amélioré chacune de mes présentations. Merci aussi d'avoir supporté la distance, les allers-retours incessants, les départs le lundi matin, les stores à moitié ouverts, mes tentatives au violon et à l'ocarina, et pour les voyages inoubliables que nous avons faits, ou même juste tous les moments passés ensemble qui m'ont permis de ne pas me transformer en zombie obsédé par ses protéines. Je n'aurais jamais aussi bien vécu cette période si tu n'avais pas été à mes côtés, et je n'aurais sûrement pas été aussi zen ; et même si tu n'as pas eu l'air de le réaliser, tu as participé au-delà de ce que tu peux imaginer à la réussite de cette thèse. Grâce à toi, ce n'est pas avec nostalgie que je la termine, mais plutôt avec impatience !

*Pour finir, pour ceux qui m'ont vu travailler les écouteurs soudés aux oreilles, voici une petite sélection musicale des titres qui m'ont accompagnés au cours de mes purifs, et qui reflète les différents états d'esprits par lesquels je suis passé :*

*Walk like an Egyptian (The Bangles), L'apprenti sorcier (Paul Dukas), No stress (Laurent Wolf), All right (Supergrass) Don't stop me now (Queen), Just my imagination (The Cranberries), Turn Around (Bonnie Tyler), Don't let me be misunderstood (Santa Esmeralda), Another one bites the dust (Queen), Try Again (Aaliyah), Don't speak (No Doubt), Relax, take it easy (Mika), Shut up and let me go! (Ting Tings) ; Hakuna Matata (Le Roi Lion), Teardrop (Massive Attack); Don't worry, be happy (Jimmy Cliff); You know that I'm no good (Amy Whine House); Tous les cris les SOS (Daniel Balavoine); Leave! (VV Brown); Le p'tit train (Rita Mitsuko); The Sweet Escape (Gwen Stefanie); Sunday Morning (The Velvet Underground); Osez, Joséphine (Alain Bashung), Sunday girl (Blondie); Stand by me (Ben E. King) ; Je ne veux pas travailler (Pink Martini); Résiste (France Gall); All night long (Lionel Richie); Zombie (The Cranberries); Suddenly I see (KT Tunstall); Smile (Lily Allen); I can see clearly now (Jimmy Cliff); Shiny Happy people (R.E.M.); Run the show! (Kat Deluna); Doctor Beat (Gloria Estefan), Party time (Gloria Estefan); Free (Stevie Wonder)*



# Table des Matières

Liste des Figures.....	1
Liste des Abréviations.....	3
Introduction.....	5
Préface - De la contingence d'un projet de thèse.....	7
Chapitre I - A la découverte de nos cousines, les Archées.....	11
A - Retour historique sur la place des micro-organismes dans le monde vivant.....	11
B - Jamais deux sans trois - Découverte d'un troisième domaine du vivant.....	19
C - Rencontre du troisième type - Aperçu général des Archées :.....	23
1 - Certains l'aiment chaud - Diversité écologique des Archées.....	23
2 - Une apparence trompeuse - les Archées, sosies des Bactéries ?.....	25
3 - Des cellules pas comme les autres - Spécificité des Archées.....	27
4 - Comment comparer Archées, Eucaryotes et Bactéries : notions de génomique comparative.....	31
5 - L'habit ne fait pas le moine - L'étonnante proximité entre Archées et Eucaryotes au niveau moléculaire.....	37
D - La génomique comparative, une fenêtre sur le passé :.....	39
1 - Le cœur des génomes - Les gènes universels :.....	41
2 - Jamais sans mon gène ! - Les gènes essentiels :.....	43
3 - Ce n'est pas la taille qui compte... Les génomes réduits :.....	45
Chapitre II - De la Traduction à t <sup>6</sup> A - Voyage au centre de l'ARNt.....	51
A- Aperçu général de la traduction.....	51
1- Un ballet moléculaire universel.....	51
2 - L'ARNt.....	57
B- Les modifications d'ARNt.....	59
1 - Les nucléotides rares, pas si rares que ça !.....	59
2 - Des adeptes du tuning... moléculaire.....	61
3 - Les enzymes de modification : un défi pour les biochimistes.....	63
4 - Rôle des nucléosides modifiés.....	67
5 - Quand la base 34 vacille, le code génétique dégénère.....	69
6 - Dangling Queen : la base 37.....	73
C – Le nucléoside modifié t <sup>6</sup> A.....	79
1 - Une modification universelle - découverte de t <sup>6</sup> A.....	79
2 - Caractérisation du rôle biologique de t <sup>6</sup> A.....	81
3 - Caractérisation de la biosynthèse de t <sup>6</sup> A.....	85
Chapitre III - La Quête de la t <sup>6</sup> A synthase.....	95
A - Sua5/YrdC (COG0009) : première pièce du puzzle de la t <sup>6</sup> A synthase.....	95
1 - Une protéine impliquée dans la traduction ?.....	95
2 - Une famille universelle... mais hétérogène :.....	99
3 - Un rôle dans la biosynthèse de t <sup>6</sup> A :.....	103
4 - Les carbamoyl-transférases HypF/NodU/CmcH/TobZ : la clé du puzzle.....	107
B - Kae1/YgjD (COG0533) : la seconde pièce du puzzle de la t <sup>6</sup> A synthétase.....	109
1 - YgjD/Kae1, une famille universelle.....	109
2 - Propriétés structurales de la famille YgjD/Kae1 :.....	113
3 - Une protéine impliquée dans la stabilité et la dynamique du génome?.....	115
4 - Kae1, l'autre enzyme de biosynthèse de t <sup>6</sup> A !.....	117

5 - Universalité du rôle de la famille Kae1/YgjD dans la biosynthèse de t <sup>6</sup> A .....	119
6 - Un puzzle complet ? .....	123
<b>C – Du côté des Bactéries : les partenaires de YgjD .....</b>	<b>123</b>
1 - YeaZ : un air de famille .....	123
2 - YjeE : un interrupteur moléculaire ? .....	127
3 - Ménage à trois.....	129
4 - La fin du mystère... chez les Bactéries ! .....	131
5 - Les protéines TobZ et HypF: une mine d'information!.....	135
<b>D - Du côté des Archées et des Eucaryotes : les partenaires de Kae1 .....</b>	<b>141</b>
1- L'intrigante Bud32, une kinase atypique .....	141
2 - Propriétés structurale et biochimique du complexe KEOPS.....	149
3 - KEOPS : un rôle biologique mystérieux mais fondamental .....	157
4 - KEOPS, impliqué dans la biosynthèse de t <sup>6</sup> A ? .....	159
<b>Objectifs .....</b>	<b>161</b>
<b>PROBLEMATIQUE .....</b>	<b>161</b>
<b>CHOIX DE L'APPROCHE .....</b>	<b>161</b>
<b>CHOIX DES MODELES D'ETUDES .....</b>	<b>163</b>
<b>DESCRIPTION DU CONTENU .....</b>	<b>165</b>

## Résultats et Discussion .....167

### Partie I - Sua5 et KEOPS constituent le Système Enzymatique de Biosynthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes.....169

<b>Résumé des résultats</b>	
<b>principaux.....</b>	<b>171</b>
<b>Article - Perrochia et al. 2013.....</b>	<b>175</b>
<b>Résultats complémentaires.....</b>	<b>177</b>
Rôle du domaine C-terminale de Sua5.....	177
Une interaction directe entre Sua5 et KEOPS? .....	179
Suivi de l'activité d'hydrolyse d'ATP par Sua5 et KEOPS.....	181
Origine de la synthèse de novo d'ADP radioactif.....	183
Effet du manganèse sur la machinerie de t <sup>6</sup> A : .....	185
Rôle de l'hydrolyse de l'ATP sur la fixation aux ARNt: .....	189
Modalités de l'hydrolyse de l'ATP par Sua5 et KEOPS : étude enzymatique préliminaire... ..	191
<b>Discussion complémentaire.....</b>	<b>197</b>
Le domaine N-terminal de Sua5, élément régulateur du domaine YrdC.....	197
Sua5 et KEOPS, des composants indépendants ?.....	199
Origine et rôle de l'activité ATPase de KEOPS.....	199
KEOPS produit de l'ADP à partir d'AMP et de Pi.....	200
Le manganèse, un cofacteur impliqué à plusieurs niveaux ? .....	201
La détermination des paramètres enzymatiques: vers la compréhension du mécanisme réactionnel.....	203
<b>Conclusion.....</b>	<b>203</b>

### Partie II - Etude du Rôle des Sous-unités de KEOPS.....205

<b>Résultats.....</b>	<b>207</b>
Purification des protéines isolées Pcc1, Kae1, Bud32, Cgi121.....	207
Purification des sous-complexes binaires et ternaires.....	211
Reconstitution physique du complexe KEOPS.....	215
Les 4 sous-unités sont requises pour la biosynthèse optimale de t <sup>6</sup> A.....	217
Le couple P-K est le principal élément de fixation aux ARNt.....	219
Identification des éléments responsables de l'activité ATPase.....	223

<b>Discussion</b>	<b>231</b>
<i>Stabilité des sous-unités et des sous complexes de KEOPS</i>	231
<i>Rôle des différentes sous-unités</i>	232
<i>Kae1 est lieu de transfert de l'AMP-TC et le centre de fixation de l'ARNt</i>	232
<i>L'activité ATPase de Bud32 traduit un mécanisme d'adaptation induite de conformation qui module l'activité de Kae1 dans la biosynthèse de t<sup>6</sup>A</i>	234
<i>La protéine Cgi121, chaperone moléculaire impliquée dans la stabilité de KEOPS ?</i>	237
<i>La protéine Pcc1, facteur de dimérisation et de fixation aux ARNt</i>	238
<i>La fixation in vivo aux ARNt nécessite l'ensemble du complexe KEOPS et permettrait d'étudier des modalités de reconnaissance</i>	240
<i>Comparaison des données biochimiques avec les données génétiques disponibles sur le rôle des sous-unités dans la biosynthèse de t<sup>6</sup>A</i>	240
<b>Conclusion</b>	<b>241</b>

## Partie III - Etude de la Machinerie Mitochondriale de Biosynthèse de t<sup>6</sup>A.....243

<b>Introduction</b>	<b>245</b>
<b>Résultats</b>	<b>249</b>
<i>Purification de Qri7</i>	249
<i>Qri7 et Sua5 de S.cerevisiae catalysent ensemble la synthèse de t<sup>6</sup>A</i>	251
<i>Qri7 et YrdC forment un système hétérologue fonctionnel de biosynthèse de t<sup>6</sup>A</i>	253
<i>Qri7 peut utiliser comme substrat des ARNt de différentes espèces</i>	253
<i>Qri7 n'utilise pas l'énergie d'hydrolyse de l'ATP pour la biosynthèse de t<sup>6</sup>A</i>	255
<i>Qri7 présente une faible affinité pour les ARNt en retard sur gel, comparable à YgjD</i>	255
<b>Discussion</b>	<b>257</b>
<i>Qri7 et Sua5 constituent la machinerie de biosynthèse de t<sup>6</sup>A chez la mitochondrie</i>	257
<i>La mitochondrie : un modèle simplifié pour l'étude du rôle in vivo de t<sup>6</sup>A</i>	258
<i>Qri7 représente un modèle simplifié pour l'étude de la machinerie de la synthèse de t<sup>6</sup>A</i>	259
<i>Qri7, une version de YgjD améliorée</i>	261
<i>La fin des débats sur la complémentarité entre YgjD et Qri7 chez E.coli ?</i>	262
<i>Une machinerie de synthèse de t<sup>6</sup>A chez les plastides</i>	263
<i>Conclusion</i>	264

## Conclusion générale et perspectives.....265

<b>Rappel de l'ensemble des résultats</b>	<b>267</b>
<b>Scénario de l'histoire évolutive de la machinerie de synthèse de t<sup>6</sup>A .</b>	<b>269</b>
<i>Un système universel, mais une évolution convergente</i>	269
<i>« La simplicité est la sophistication extrême » (Leonard de Vinci) – le cas de la mitochondrie</i>	271
<i>Chez les Bactéries : d'un homodimère à un hétérodimère</i>	272
<i>Du côté des Archées et des Eucaryotes...</i>	273
<b>La biosynthèse de t<sup>6</sup>A, entre universalité et variabilité</b>	<b>276</b>
<i>Le premier composant Sua5/YrdC, une activité universelle, une variabilité structurale intrigante</i>	276
<i>Le second composant Kae1/YgjD : une activité universelle de transfert du TC sur l'ARNt</i>	277
<i>KEOPS et DEZ : des analogues fonctionnels... mais des fonctionnements différents</i>	279
<i>Des déterminants universels pour la reconnaissance de l'ARNt</i>	280
<i>... mais une exception intrigante</i>	281
<i>Les composants de la t<sup>6</sup>A synthase, impliqués dans d'autres fonctions biologiques ?</i>	283
<b>Epilogue : une histoire de nom</b>	<b>285</b>

<b>Bibliographie.....</b>	<b>287</b>
---------------------------	------------

<b>Matériel et Méthodes.....</b>	<b>299</b>
----------------------------------	------------

<i>Clonage par synthèse de gène (E, Sua5ΔC, Sua5ΔN, KEOPS, PKB) .....</i>	<i>301</i>
<i>Clonages par PCR (YgjD, YeaZ, YrdC, Qri7, Pcc1, Kae1, Bud32, Cgi121, PK, KB, BC, KBC, PKB) .....</i>	<i>302</i>
<i>Production de protéines recombinantes par sur-expression dans Escherichia coli.....</i>	<i>303</i>
<i>Purification des protéines recombinantes par chromatographie d'affinité sur résine NiNTA.....</i>	<i>304</i>
<i>Purification des protéines recombinantes portant une étiquette StrepTag II.....</i>	<i>306</i>
<i>Purification des protéines recombinantes par gel filtration .....</i>	<i>306</i>
<i>Analyse du spectre d'absorption et détermination de la concentration de protéine.....</i>	<i>307</i>
<i>Détermination de la concentration protéique par quantification au réactif de Bradford .....</i>	<i>307</i>
<i>Analyse SDS PAGE des protéines .....</i>	<i>307</i>
<i>Analyse des acides nucléiques liés aux protéines en gel d'agarose .....</i>	<i>308</i>
<i>Construction des plasmides de sur-expression des ARNt .....</i>	<i>308</i>
<i>Production d'ARNt par sur-expression dans Escherichia coli.....</i>	<i>309</i>
<i>Purification des ARNt par la méthode "crush and soak" .....</i>	<i>309</i>
<i>Mesure de l'incorporation in vitro de la C14 L-thréonine.....</i>	<i>310</i>
<i>Extraction des ARNt après réaction d'incorporation in vitro: .....</i>	<i>311</i>
<i>Digestion des ARNt en mononucléotides et séparation par chromatographie sur couche mince en 2D dimensions.....</i>	<i>312</i>
<i>Activité ATPase en chromatographie sur couche mince.....</i>	<i>313</i>
<i>Test de phosphorylation en SDS PAGE.....</i>	<i>314</i>
<i>Test d'interaction par gel filtration .....</i>	<i>314</i>
<i>Retard sur gel d'acrylamide en condition native .....</i>	<i>315</i>

<b>Annexes.....</b>	<b>319</b>
---------------------	------------

<b>Annexe 1: A Guide to Understanding Journal Phraseology.....</b>	<b>321</b>
<b>Annexe 2 : Table des oligonucléotides.....</b>	<b>322</b>
<b>Annexe 3 : Séquence et propriétés des protéines purifiées.....</b>	<b>325</b>
<b>Annexe 4 : Composition des tampons et protocoles.....</b>	<b>328</b>

<b>Résumé/Summary.....</b>	<b>331</b>
----------------------------	------------

# Liste des Figures

## Introduction

Figure 1	Extraits de l'essai « Histoire naturelle des Animaux sans vertèbres » de J.B. Lamarck, publié en 1837.....	12
Figure 2	Le premier arbre du vivant d'E. Haeckle publié en 1866.....	14
Figure 3	Les 5 règnes du vivant selon Whittaker (1969), adapté.....	16
Figure 4	Extraits de l'article de C. Woese, « There must be a prokaryote somewhere », publié en 1994.....	20
Figure 5	Aperçu de la diversité morphologique des Archées.....	24
Figure 6	La membrane des Archées, une spécificité unique dans le monde vivant.....	26
Figure 7	Comparaison entre le flagelle bactérien, le « flagelle » archéen et le pili de type IV des Bactéries.....	28
Figure 8	Illustration des notions d'orthologie et de paralogie.....	30
Figure 9	Répertoire original des Groupes de Gènes Orthologues (COG) établi en 1997.....	32
Figure 10	Comparaison structurale des ARN polymérases (RNAP) issues des trois domaines du vivant.....	36
Figure 11	Exemples de scénarii simplifiés expliquant les liens et l'évolution des trois domaines du vivant.....	38
Figure 12	Distribution des gènes conservés entre différentes espèces.....	40
Figure 13	Illustration du biais induit par la comparaison des COG entre 7 organismes.....	42
Figure 14	Etude des gènes essentiels.....	44
Figure 15	Représentation du génome mitochondrial humain, un exemple de génome très réduit.....	46
Figure 16	Le grand théâtre de la Traduction et ses principaux acteurs.....	50
Figure 17	Une danse à trois partenaires.....	52
Figure 18	Représentation de la structure secondaire « en feuille de trèfle » des ARNt.....	56
Figure 19	Représentation de la structure tridimensionnelle en « L » classique d'un ARNt.....	56
Figure 20	Distribution des nucléosides modifiés.....	58
Figure 21	Composition chimique des différents nucléosides modifiés.....	60
Figure 22	Exemples de modifications chimiques à l'origine de certains nucléosides modifiés.....	62
Figure 23	Exemple de modifications catalysées par des enzymes de classe I et de classe II .....	64
Figure 24	Type et localisation des nucléosides modifiés dans la boucle anticodon des ARNt.....	66
Figure 25	Le Code Génétique.....	68
Figure 26	L'hypothèse du Wobbling et sa version « modifiée ».....	70
Figure 27	Structure chimique du nucléoside N <sup>6</sup> -thréonylcarbamoyladénosine (t <sup>6</sup> A) et de ses variants.....	78
Figure 28	t <sup>6</sup> A et ses variants sont présents sur les ARNt reconnaissant les codons de la forme ANN.....	80
Figure 29	Illustration du rôle structural de t <sup>6</sup> A au niveau de la boucle anticodon de l'ARNt.....	82
Figure 30	Précurseurs de t <sup>6</sup> A identifiés en 1975.....	84
Figure 31	Technique de micro-injection d'ARNt dans des ovocytes de Xénope pour l'étude des modifications d'ARNt <i>in cellulo</i> .....	84
Figure 32	Représentation des déterminants moléculaires nécessaires à la machinerie de biosynthèse de t <sup>6</sup> A chez <i>Xenopus laevis</i> .....	86
Figure 33	L'ARNt <sup>Met</sup> initiateur révèle des différences entre les machineries de synthèse t6A chez les trois domaines du vivant.....	88
Figure 34	Nomenclature des différentes protéines impliquées dans la voie de biosynthèse de t <sup>6</sup> A .....	92
Figure 35	Comparaison des structures des protéines YrdC d' <i>Escherichia coli</i> de Sua5 de <i>Sulfolobus tokodaii</i> .....	96
Figure 36	Représentation des orthologues de YrdC/Sua5 représentatifs de différents groupes d'organismes.....	98
Figure 37	Comparaison des protéines contenant les domaines structuraux YrdC et YgjD.....	106
Figure 38	Phylogénie de la famille YgjD/Kae1.....	110
Figure 39	Structure de la protéine Kae1 de <i>Pyrococcus abyssi</i> .....	112
Figure 40	Mise en évidence des différences structurales entre YeaZ et Kae1.....	122
Figure 41	Comparaison des formes homodimériques de la protéine YeaZ de <i>Thermotoga maritima</i> .....	124
Figure 42	Structure de l'hétérodimère YeaZ-YgjD de <i>Salmonella typhimurium</i> , cristallisé en présence d'ATPγS.....	126
Figure 43	Structure de la protéine YjeE d' <i>Haemophilus influenzae</i> .....	126
Figure 44	Contexte génomique des gènes <i>yjeE</i> , <i>yeaZ</i> , <i>ygdD</i> et <i>yrdC</i> chez certaines espèces de Bactéries.....	128
Figure 45	Modèles d'interaction entre les protéines YgjD, YeaZ, YjeE et YrdC.....	130

Figure 46	Modèle de biosynthèse de t <sup>6</sup> A chez les Bactéries par les protéines YgjD, YeaZ, YjeE et YrdC.....	132
Figure 47	Structure et mécanisme catalytique de la protéine TobZ, illustrant le rôle des domaines YrdC et Kae1.....	134
Figure 48	Propriétés structurales de Bud32 de <i>M.jannaschii</i> et comparaison avec d'autres kinases.....	142
Figure 49	Composition et nomenclature du complexe KEOPS.....	146
Figure 50	Architecture linéaire du complexe KEOPS chez les Archées.....	148
Figure 51	Structure de la protéine MJ1130 de <i>M.jannaschii</i> , et mise en évidence des interactions entre les domaines Kae1/Bud32....	150
Figure 52	Structure de la protéine Cgi121 de <i>M.jannaschii</i> .....	152
Figure 53	Structure de la protéine Pcc1 de <i>T. acidophilum</i> , comparaison avec le domaine KH et structure du complexe Kae1/Pcc1.....	154

## Résultats et Discussion

Figure I-1	Etude des domaines N et C terminaux de Sua5.....	176
Figure I-2	Retard sur gel des ARNt par KEOPS en présence de Sua5.....	178
Figure I-3	Suivi parallèle de l'activité ATPase par l'utilisation d'ATP <sup>[αP32]</sup> et d'ATP <sup>[γP32]</sup> .....	180
Figure I-4	Etude de la formation de novo d'ADP radioactif.....	182
Figure I-5	Effet du MnCl <sub>2</sub> (manganèse) sur les activités de KEOPS et Sua5.....	184
Figure I-6	Fixation de l'ARNt par KEOPS en présence de différents nucléotides.....	188
Figure I-7	Modalités de l'activité ATPase de Sua5.....	190
Figure I-8	Modalités de l'activité ATPase de KEOPS.....	194
Figure II-1	Expression et purification des sous unités Pcc1, Cgi121 et Bud32 de <i>Pyrococcus abyssi</i> .....	206
Figure II-2	Purification de la protéine Cgi121 par filtration sur gel.....	208
Figure II-3	Purification des sous-complexes de KEOPS.....	210
Figure II-4	Analyse par filtration sur gel de la taille apparente des sous-complexes de KEOPS.....	212
Figure II-5	Test d'interaction entre les sous-complexes de KEOPS par gel filtration.....	214
Figure II-6	Activité t <sup>6</sup> A des différents sous-complexes de KEOPS.....	216
Figure II-7	Propriétés de fixation à l'ARNt par les sous-complexes de KEOPS en retard du gel.....	218
Figure II-8	Activité ATPase des sous complexes de KEOPS.....	222
Figure II-9	Activité ATPase des sous complexes mutants.....	224
Figure II-10	Activité phosphotransférase de Bud32 dans différentes conditions.....	226
Figure II-11	Activité t <sup>6</sup> A des sous-complexes mutants de KEOPS.....	226
Figure III-1	Purification de la protéine Qri7 de <i>Saccharomyces cerevisiae</i> .....	248
Figure III-2	Activité t <sup>6</sup> A de la protéine Qri7 avec Sua5 et YrdC.....	250
Figure III-3	Suivi de l'activité ATPase de Qri7 et Sua5 de <i>Saccharomyces cerevisiae</i> .....	252
Figure III-4	Analyse de la fixation d'ARNt par Qri7, YgjD en retard sur gel de polyacrylamide en condition native.....	254

## Conclusion

Figure C-1	Représentation schématique des différentes machineries de synthèse de t <sup>6</sup> A des espèces modèles <i>P. abyssi</i> (Archée), <i>S. cerevisiae</i> (Eucaryote) et <i>E. coli</i> (Bactérie).....	268
Figure C-2	Scénario hypothétique de l'évolution de la machinerie de biosynthèse de t <sup>6</sup> A dans les trois domaines du vivant.....	270



# Liste des Abréviations

ADN	=	Acide DésoxyriboNucléique
ADP	=	Adénosine Di-Phosphate
AMP	=	Adénosine Mono-Phosphate
AMPPNP	=	Adénosine diphosphate-Nitro-Phosphate
ARN	=	Acide RiboNucléique
ARNm	=	ARN messenger
ARNr	=	ARN ribosomique
ARNt	=	ARN de transfert
ATP	=	Adénosine Tri-Phosphate
BET	=	Bromure d'Ethidium
BSA	=	Bovine Serum Albumin ( <i>Albumine de Sérum de Boeuf</i> )
<sup>14</sup> C	=	Carbone 14
CCM	=	Chromatographie sur Couche Mince
COG	=	Cluster of Orthologous Groups of genes
CPM	=	Coup Par Minute
DTT	=	DiThioThréiol
EDTA	=	EthyleneDiamineTetraAcetic acide
EKC	=	Endopeptidase Kinase Chromatin associated
FPLC	=	Fast Protein Liquid Chromatography
G	=	Guanosine
HPLC	=	High Performance Liquid Chromatography ( <i>chromatographie en phase liquide à haute performance</i> )
IPTG	=	Isopropyl β-D-1-thiogalactopyranoside
kb	=	kilobase
kDa	=	kiloDalton
KEOPS	=	Kinase, Endopeptidase and Other Proteins of Small size
LB	=	Lysogeny Broth*
LUCA	=	Last Universal Common Ancestor ( <i>Dernier ancêtre commun universel</i> )
Mb	=	Megabase
MCS	=	Multiple Cloning Site ( <i>site de clonage multiple</i> )
ORF	=	Open Reading Frame ( <i>cadre ouverte de lecture</i> )
PCR	=	Polymerase Chain Reaction ( <i>Réaction de polymérisation en chaîne</i> )
Pi	=	Phosphate inorganique
PPi	=	PyroPhosphate inorganique
RBS	=	Ribosome Binding Site ( <i>site de fixation du ribosome</i> )
SDS PAGE	=	Sodium Dodécylsulfate PolyAcrylamide Gel Electrophoresis
t <sup>6</sup> A	=	N <sup>6</sup> -thréonyl-carbamoyl-Adénosine
TB	=	Terrific Broth
TCA	=	TriChloracetic Acid
TC-AMP	=	ThreonylCarbamoyl-Adénosine MonoPhosphate
TCL	=	Thin Layer Chromatography ( <i>Chromatographie sur couche mince ou CCM</i> )
Tris	=	Tris(hydroxyméthyl)-aminométhane

\*Giuseppe Bertani, *Lysogeny at Mid-Twentieth Century: P1, P2, and Other Experimental Systems*, Journal of Bacteriology, February 2004, "My first paper on lysogeny, describing the modified single-burst experiment and the isolation of P1, P2, and P3, also contained the formula of the LB medium which I had concocted in order to optimize *Shigella* growth and plaque formation. Its use has since become very popular. The acronym has been variously interpreted, perhaps flatteringly, but incorrectly, as Luria broth, Lennox broth, or Luria-Bertani medium. For the historical record, the abbreviation LB was intended to stand for "lysogeny broth."



A large, stylized illustration of various microscopic organisms, including bacteria, viruses, and fungi, arranged in a circular pattern around the central text. The organisms are depicted in black and white, with some showing flagella, spores, and complex cellular structures.

# INTRODUCTION



# Préface- De la contingence d'un projet de thèse

*« Chaque fois que l'on redéroule le film de la vie, l'évolution prend une voie différente de celle que nous connaissons. On s'aperçoit que chaque nouvelle voie empruntée est tout aussi interprétable, tout aussi explicable a posteriori que celle qui a été réellement suivie et que nous connaissons. Mais la diversité des itinéraires possibles montre à l'évidence que les résultats finaux ne peuvent être prédits au départ. Chacune des étapes a ses propres causes, mais on ne peut dire quels états finaux seront réellement atteints ; et aucun de ceux-ci ne sera à nouveau obtenu lorsqu'on redéroulera le film, parce que chacune des nouvelles voies se réalise par l'enchaînement de milliers d'étapes imprévisibles. Changez faiblement les événements initiaux, si faiblement que cela peut paraître sur le moment n'avoir qu'une minime importance, et l'évolution se déroulera selon une direction toute différente. Cette alternative ne représente ni plus, ni moins que l'essence de l'histoire. Elle a pour nom contingence - et la contingence est une chose en soi, et non la combinaison du déterminisme et du hasard. »*

**Stephen Jay Gould, extrait de *La vie est Belle***

En 2008, je suis tombé presque par hasard à la FNAC sur un livre dont le titre m'a immédiatement attiré : « Microbes de l'Enfer ». Il était question d'Archées, d'organismes hyperthermophiles et de l'histoire passionnante des chercheurs partis à leur recherche. J'étais déjà tombé amoureux des Bactéries (grâce également à un autre livre passionnant, « *Le Monde Bactérien* » de Lynn Margulis), et je connaissais plus ou moins l'existence des Archées, en tant que groupe d'organismes singuliers et extrêmophiles. J'ai compris alors qu'ils représentaient un monde bien à part, rempli de surprises et de découvertes étonnantes.

Quelques mois plus tard, durant le printemps 2009, en pleine période de révision du concours de l'Agrégation, je devais chercher un stage de recherche dans le cadre de ma deuxième année de Master. J'avais alors une idée assez précise de la thématique vers laquelle je souhaitais m'orienter : l'étude des organismes extrêmophiles, représentant pour

moi l'accès à une thématique sous-jacente qui me passionnait (et qui me passionne toujours), l'exobiologie. Je n'avais pourtant aucune idée du laboratoire vers lequel me tourner. Après quelques recherches infructueuses, j'ai commencé à me dire que l'idée de travailler dans ce domaine allait être difficile à concrétiser. C'est à ce moment qu'intervient la première part de contingence de mon projet de recherche. Par chance, j'avais laissé trainer « Microbe de l'Enfer » sur ma table de nuit. Je me suis alors rappelé en le voyant que l'auteur, Patrick Forterre, était un chercheur, français de surcroît, et que sa thématique de recherche correspondait exactement à ce que je recherchais. J'ai alors décidé de lui envoyer un e-mail, pour lui demander s'il serait possible de venir faire un stage dans son équipe. Nous nous sommes donc rencontrés en Juin 2009, à Orsay. Si je n'étais pas tombé sur ce livre quelques mois plus tôt, ou si je l'avais rangé au fin fond de ma bibliothèque, il est probable que les choses aient pris une tournure différente.

Lors de notre entrevue, Patrick m'a alors proposé trois sujets de stage. D'une part, la thématique de prédilection du laboratoire : l'étude des topoisomérases. D'autre part, la mise en place d'un système de réplication *in vitro* à partir de protéines d'Archées. Enfin, dernier choix possible, l'étude d'une « mystérieuse protéine universelle » (*dixit* Patrick), faisant partie d'un complexe appelé KEOPS, dont la fonction était encore inconnue. Ce dernier choix m'a immédiatement attiré, principalement à cause du mot « mystérieux » mais aussi probablement dans mon inconscient à cause du nom « KEOPS », étant amateur d'égyptologie. Si Patrick avait employé l'autre appellation de KEOPS, à savoir EKC, et si il n'avait pas employé le mot « mystérieux », j'aurais peut-être choisi l'une des deux autres propositions et ce manuscrit n'aurait pas eu le même contenu.

J'ai donc rejoint l'équipe BMGE d'Orsay en Septembre 2009, et j'ai alors appris que l'origine de ce projet KEOPS avait lui aussi sa part de contingence. A l'origine, Patrick et son équipe cherchaient à comprendre les mécanismes de la réplication de l'ADN chez les Archées hyperthermophiles. Dans le début des années 2000, ils avaient identifié les protéines impliquées dans cette réplication, qui s'avèrent être des homologues de celles des Eucaryotes. Ils se mirent alors à chercher si, comme chez les Eucaryotes, ces protéines étaient régulées par phosphorylation. En 2003, Edwige Madec, post-doctorante, initia donc ce projet en recherchant les kinases impliquées dans la réplication de l'ADN chez *Pyrococcus abyssi*. Plusieurs candidats furent identifiés, parmi lesquels se trouvait la protéine PAB1047,

appelée par la suite Bud32. Or, il s'avéra que cette protéine était souvent fusionnée chez les Archées à une autre protéine de fonction inconnue, appelée Kae1. Patrick réalisa que les Archées, les Eucaryotes et même les Bactéries possédaient un orthologue de Kae1. Une protéine universelle, présente dans les trois domaines du vivant, probablement chez LUCA (le dernier ancêtre commun universel du vivant), et dont la fonction restait encore inconnue ?! Le mystère était trop attirant, et Arnaud Hecker, qui prit la relève du projet, se concentra alors sur l'étude de cette « mystérieuse » protéine, aux côtés de Jacques Oberto. Encore une fois, si Bud32 n'avait pas été l'un des candidats identifiés à l'origine, et si Patrick n'avait pas eu un amour invétéré pour les processus universels du vivant, il est possible que je ne sois jamais amené à travailler sur Kae1.

J'ai donc participé durant l'année 2009/2010, encadré et formé par Arnaud, et en collaboration avec les équipes de H. Van Tilbeurgh et D. Libri, à la caractérisation du complexe KEOPS auquel Kae1 et Bud32 appartiennent, afin de comprendre quelle était sa fonction chez les Archées et les Eucaryotes. Les pistes étaient nombreuses, et les données semblaient constituer un grand puzzle dont il manquait encore une pièce maîtresse. Au cours du mois de Mai 2010, Arnaud « tomba » sur le résumé d'un poster qui impliquait YgjD, l'homologue bactérien de Kae1, dans la voie de biosynthèse d'une modification universelle d'ARNt, t<sup>6</sup>A. Nous avons alors rencontré Valérie De Crécy-Lagard, qui nous expliqua que son équipe était sur le point de publier cette découverte. La pièce manquante du puzzle semblait donc avoir été trouvée : KEOPS pourrait jouer un rôle dans cette voie de biosynthèse.

J'ai alors été confronté à un choix crucial pour mon projet de thèse : continuer le travail sur KEOPS, au risque de m'engager dans un nouveau domaine qui était totalement étranger au laboratoire, et me confronter à une compétition directe, ou m'orienter vers un autre sujet de recherche moins risqué. Encore une fois, si j'avais fait un autre choix, le contenu de cette thèse aurait probablement été radicalement différent. La biologie des ARNt et de leurs modifications étaient pour moi un monde totalement inconnu, obscur et, je dois avouer, assez effrayant ; mais après avoir longuement hésité, j'ai décidé que le challenge était trop passionnant pour l'abandonner. Tamara Basta, qui a pris la relève d'Arnaud dans mon encadrement, m'a alors accompagné dans ce projet, rejointe par Estelle Crozat puis Dorian Guetta. Nous nous sommes donc engagés dans la caractérisation de la machinerie de synthèse de t<sup>6</sup>A chez les Archées et chez les Eucaryotes, ce qui m'a permis de

réaliser que le monde des modifications d'ARNt, loin d'être effrayant, était en réalité un univers passionnant.

Voilà la suite d'évènements qui expliquent comment j'ai fini par travailler sur un sujet auquel je n'aurais jamais pensé quelques années plus tôt, et pourquoi l'équipe de Patrick Forterre, spécialiste de la réplication de l'ADN à haute température, participe maintenant à des congrès sur la biologie des ARNt. L'introduction qui suit reprend certains des éléments dont je viens de parler, mais assemblés d'une manière différente, qui ne reflète pas nécessairement la réalité historique contingente de ce projet. C'est la raison pour laquelle je souhaitais exposer au préalable ce bref historique.

Ce travail de thèse s'articule autour de deux points de vue imbriqués: un point de vue biochimique et un point de vue évolutif sous-jacent. J'ai donc tenu dans une première partie d'introduction à revenir sur un historique de la classification des organismes, afin de présenter la découverte des Archées et ses conséquences dans la vision actuelle du monde vivant. J'ai choisi un ton délibérément militant, en hommage à Carl Woese. Je présenterai ensuite quelques aspects caractéristiques des Archées, puis j'aborderai certaines notions de génomique comparative. A l'issue de cette partie, je montrerai en quoi la traduction est l'un des processus les plus anciens et les mieux conservés du monde vivant.

Dans une deuxième partie, j'aborderai les grands principes concernant la traduction et la structure des ARNt, avant de présenter les modifications d'ARNt et leur importance, pour aboutir à la modification universelle  $t^{\text{f}}\text{A}$  dont je détaillerai les connaissances actuelles.

Enfin, dans une troisième partie, je présenterai les travaux concernant les protéines liées à la biosynthèse de  $t^{\text{f}}\text{A}$ , et qui sont l'objet de cette thèse. Les protéines seront abordées en fonction de l'ordre chronologique dans lequel elles ont été impliquées dans cette voie métabolique. Je parlerai ainsi dans un premier temps de la famille de protéine universelle YrdC/Sua5, puis de la famille universelle Kae1/YgjD. Je présenterai par la suite les partenaires de la protéine YgjD chez les Bactéries et la machinerie bactérienne de synthèse de  $t^{\text{f}}\text{A}$  qu'elles constituent, pour enfin présenter les partenaires Archées/Eucaryotes de la protéine Kae1, à savoir les membres du complexe KEOPS, et aboutir finalement à la problématique de mon travail de thèse.



# Chapitre I - A la découverte de nos cousines, les Archées

## A - Retour historique sur la place des micro-organismes dans le monde vivant

Nous, êtres humains, *Homo sapiens*, appartenons à l'ordre des Primates, à la classe des Mammifères, à l'embranchement des Chordés, et à l'Infra-règne des animaux Bilatériens (HAECKEL, 1866). Ce dernier terme représente sans doute l'un des aspects les plus facilement reconnaissables au premier coup d'œil : deux mains, deux pieds, deux yeux, deux oreilles... Notre cerveau aussi, sous son apparente complexité, est construit et fonctionne selon ce modèle d'une symétrie bilatérale : le cerveau « gauche », analytique, le cerveau « droit », créatif. Cette nature fondamentalement bilatérienne de l'être humain explique peut être pourquoi notre logique et nos réflexions aiment à concevoir, classer et ordonner les concepts fondamentaux sous forme de dichotomies : Bien *Versus* Mal, Ying *Versus* Yang, Masculin *Versus* Féminin, Sucré *Versus* Salé, Orient *Versus* Occident, Animaux *Versus* Plante.

Les sciences de la classification du vivant (taxonomie et systématique), de même que plus généralement la biologie, n'ont pas échappé à cette règle, et pendant deux siècles, le spectre de la dichotomie a hanté notre conception du monde vivant. Cependant, durant cette période, de minuscules organismes, appelés encore vulgairement microbes, ont attaqué, entaillé et fissuré l'édifice de la Sainte dichotomie, jusqu'à le faire s'effondrer finalement en 1977. Revenons donc dans le passé, au siècle des Lumières, époque où l'on commença à s'intéresser de manière scientifique à la classification du vivant.

### LES INFUSOIRES. (Infusoria.) (1).

Animaux microscopiques, gélatineux, transparents, polymorphes, contractiles.  
- Point de bouche distincte; aucun organe intérieur constant, déterminable; génération fissipare, subgemmipare.

### INFUSOIRES NUS.

Monade.	Gone.
Volvoce.	Cyclide.
Protée.	Paramée.
Enchélide.	Kolpode.
Vibrion.	Bursaire.

### INFUSOIRES APPENDICULÉS.

Tricode.	} Point de queue.
Kérone.	
Cercaire.	} Une queue.
Furcocerque.	

## HISTOIRE NATURELLE

DES

# ANIMAUX SANS VERTÈBRES

J. B. P. A. DE LAMARCK,

Membre de l'Institut de France, Professeur au Muséum d'Histoire Naturelle.

Troisième Édition,

REVUE ET AUGMENTÉE DE NOTES PRÉSENTANT LES FAITS NOUVEAUX DONT LA SCIENCE S'EST ENRICHIE  
JUSQU'À CE JOUR;

1837

Ces petits animaux exigent des observations microscopiques très-déliées, une patience presque sans bornes pour reconnaître les faits qu'ils nous présentent, enfin, un esprit libre ou dégagé de préjugés, afin de ne voir en eux que ce qui y est véritablement.

Lorsqu'on manque de loisirs ou de moyens pour les observer soi-même, il faut, pour s'en procurer la notion, consulter les ouvrages de Leuwenhoeck, qui en fit la découverte; d'Othon-Frédéric Muller, qui en observa un très-grand nombre et en décrivit beaucoup de genres et d'espèces; en un mot, ceux

Si on laisse quelque temps de l'eau exposée à la chaleur de l'air ou du soleil, et surtout de l'eau dans laquelle des matières animales ou végétales ont été infusées, on y voit bientôt paraître de ces infusoires; mais on ne peut en général les apercevoir qu'avec le secours du microscope.

L'existence des infusoires et l'état réel de leur organisation et de leurs facultés sont les seuls objets qui puissent nous intéresser à leur égard. Aussi ce n'est que philosophiquement et comme des objets de première importance à considérer dans l'étude de la nature, que nous devons nous en occuper.

Les infusoires nus sont des animalcules très-simples, infiniment-petits, la plupart transparents, dépourvus, au moins en apparence, d'appendices extérieurs, comme de poils, de cils, de sorte de cornes ou d'une queue, et qui ne paraissent, sous l'œil armé, que des points animés ou mouvants (1). Ces animalcules, et surtout parmi eux ceux qui ont le corps globuleux ou sphérique, offrent ce qu'il y a de plus simple dans le règne animal, c'est-à-dire, les plus faibles échelons de l'organisation.

Parmi les merveilles sans nombre que la nature offre de toutes parts à nos observations, la plus étonnante peut-être, c'est de voir la vie animale pouvoir exister dans des corps aussi frêles et aussi simples que ceux qui constituent les animaux de cette classe, et surtout de son premier ordre.

En effet, les infusoires, considérés dans ceux dont j'assigne le caractère classique, nous présentent l'organisation animale dépourvue de tout organe particulier intérieur, constant et déterminable, réduite à n'offrir qu'une masse de tissu cellulaire variée, extrêmement petite, frêle, presque sans consistance, et cependant vivante et très-irritable.

Les infusoires, même les plus imparfaits, sont donc tous véritablement des animaux, puisque de proche en proche ils sont liés les uns aux autres par des rapports évidents, et qu'ils conduisent, sans lacune, aux polypes qui sont bien reconnus pour appartenir au règne animal.

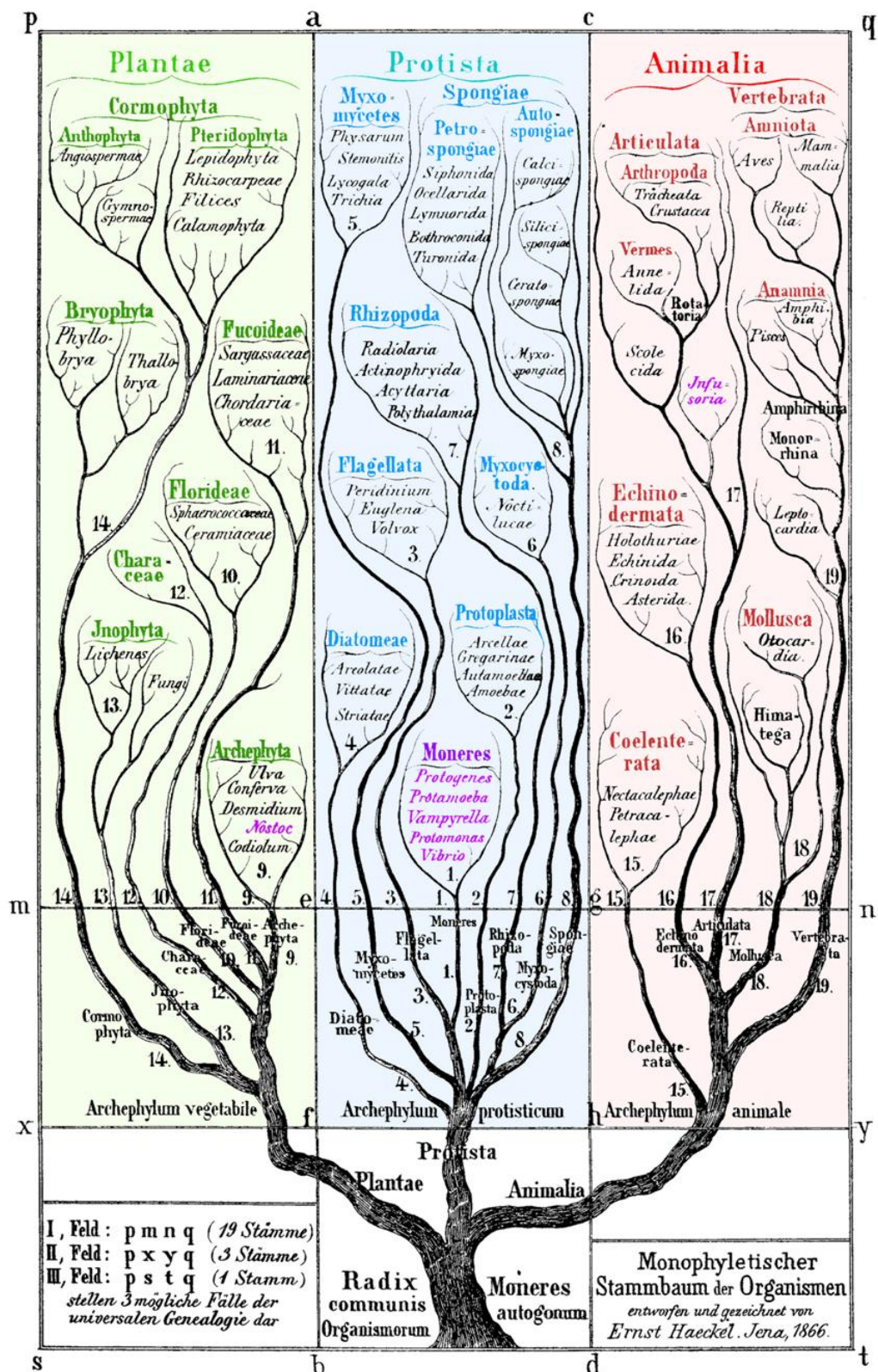
**FIGURE 1** : Extraits de l'essai « Histoire naturelle des Animaux sans vertèbres » de J.B. Lamarck, publié en 1837, illustrant la vision des micro-organismes appelés « Infusoires » au XIX<sup>ème</sup> siècle.

De 1735 à 1768, le naturaliste danois Carl Van Linné publia plusieurs éditions de *Systema Naturae*, ouvrage fondamental de classification de l'ensemble des être vivants. Il introduisit la nomenclature binomiale classique encore utilisée aujourd'hui, et classa les organismes en grands ordres, répartis au sein d'une dichotomie fondamentale : le Règne Animal *versus* Règne Végétal. Dans cette grande entreprise de classification, on peut trouver une espèce obscure dans l'ordre des *Vermes* (les Vers), placée un peu comme la poussière sous le tapis : *Chaos infusoria*. Elle regroupe un ensemble d'organismes microscopiques décrits à l'époque, appelés Infusoires, qui apparaissent « en laissant des matières animales et végétales infuser dans de l'eau ». Connus depuis les observations d'Antoni Van Leeuwenhoek (1632-1723) au siècle précédent, leur existence, leur origine et leurs caractéristiques étaient confuses. Le nom de *Chaos infusoria* fut prémonitoire : la place et la classification des micro-organismes restèrent en effet relativement chaotiques durant les 200 ans suivants.

La suggestion d'un troisième groupe du vivant, distinct des Plantes et des Animaux, les Psychodaires (« êtres ayant deux âmes », à la fois animaux et végétaux), par Jean Baptiste Bory de Saint Vincent (1778-1846) en 1824, permettant de trouver une place à ces Infusoires, ne connut pas un véritable succès. Jean-Baptiste de Lamarck (1744-1829) (*Histoire naturelle des animaux sans vertèbres*) s'essaya aussi à déterminer la nature des Infusoires, leur conférant un caractère clairement animal (**FIGURE 1**)

Au cours du XIX<sup>ème</sup> siècle, le développement des techniques de microscopie et la naissance de la microbiologie moderne (grâce notamment aux travaux de Louis Pasteur (1822-1895)) transformèrent les Infusoires flous et incompris en un monde d'une grande diversité. Un monde que l'on commença à caractériser, ordonner, maîtriser, voire combattre. Le terme germe fut utilisé en 1874 par Joseph Lister (1827-1912), puis le terme microbe fit son apparition 2 ans plus tard. Dans les débuts des années 1870, le terme bactérie (en grec pour petit bâton) fut également employé pour désigner les plus petits germes (**PACE ET AL, 2012**, revue détaillée sur l'histoire de la microbiologie)





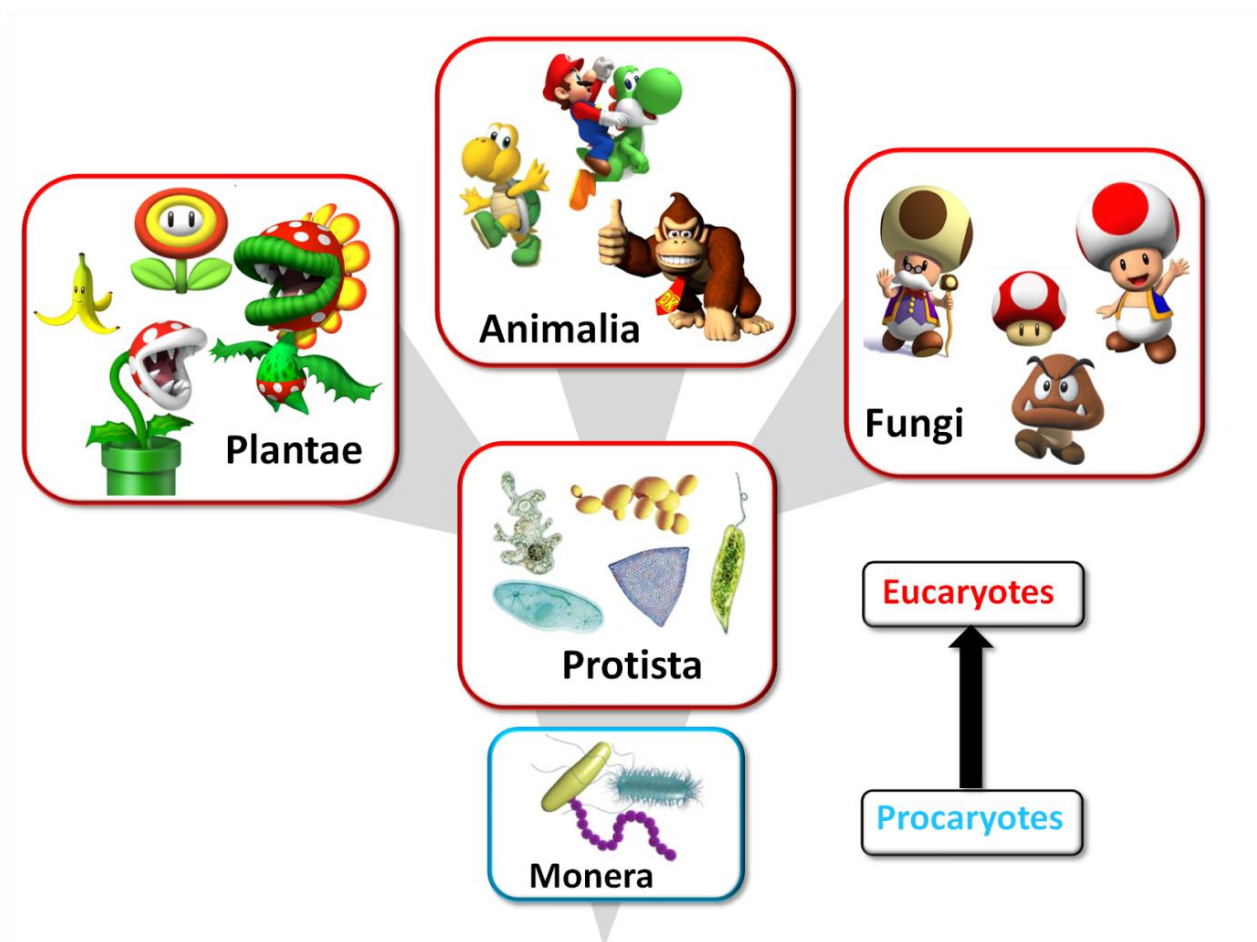
**FIGURE 2 :** Le premier arbre du vivant d'E. Haeckel publié en 1866, où apparaît un troisième domaine entre les Plantes (en vert) et les Animaux (en rouge) : les Protistes (en bleu). Les bactéries connues à l'époque ont été placées dans le groupe des Monères (en violet). Certaines bactéries ont été placées parmi les Animaux (les Infusoires) ou les Plantes (Nostoc, correspondant aux cyanobactéries).

Mais où placer ces organismes? Animaux, Végétaux? La réponse restait toujours incertaine et variable. Louis Pasteur en parlait lui-même parfois comme des plantes microscopiques, parfois comme des animalcules. En 1866, Ernst Haeckel (1834-1919), à qui l'ont doit les termes d'ontogénie, phylogénie et phylum, fut l'un des premiers à proposer une classification du vivant abordant des questions évolutives. S'inspirant des travaux contemporains et novateurs de Charles Darwin sur l'origine des espèces, il proposa un arbre phylogénique universel comprenant non pas deux mais bien trois branches. (FIGURE 2)

Entre le Règne Animal et le Règne Végétal, il introduisit un troisième grand règne, les Protistes, réunissant les micro-organismes unicellulaires et incluant le sous-groupe des Monères, organismes simples (dépourvu de structure complexe, notamment de noyau) correspondant aux Bactéries (les « algues » bleues-vertes, ou cyanobactéries, ne furent placée au sein des Monères qu'en 1904 par Haeckel) Ces Monères représentaient la transition entre le non vivant et le vivant. Mettant un terme au flou caractérisant les micro-organismes, cet arbre fut controversé et ne prit pas racine dans les esprits ; on s'accommoda finalement plus ou moins à l'idée de placer, en poussant un peu « dans les coins », certains protistes du côté animaux (protozoaires), d'autre du côté végétaux (protophytes), les Bactéries revêtant alors un caractère plutôt végétal

Ainsi, à travers presque tout le XX<sup>ème</sup> siècle, la microbiologie et la bactériologie restèrent principalement la chasse gardée des botanistes, enseignées comme une sous-discipline. Les Bactéries pratiquant la photosynthèse étaient qualifiées d'« algues », les autres de « schizomycètes » (= « végétaux unicellulaires sans chlorophylle », Petit Larousse 1906). Cette idée a laissé des reliques jusqu'à aujourd'hui dans l'inconscient collectif. Ne parle-t-on pas de flore intestinale pour désigner les micro-organismes commensaux ?

La première moitié du XX<sup>ème</sup> siècle verra cependant l'évolution des concepts, une dichotomie en remplaçant une autre. En 1962, Stanier et Van Niel écrivirent un article, « *The concept of a Bacterium* », dans le but de résoudre les problèmes concernant le concept et la définition de la bactérie une fois pour toute (STANIER, VAN NIEL, 1962).



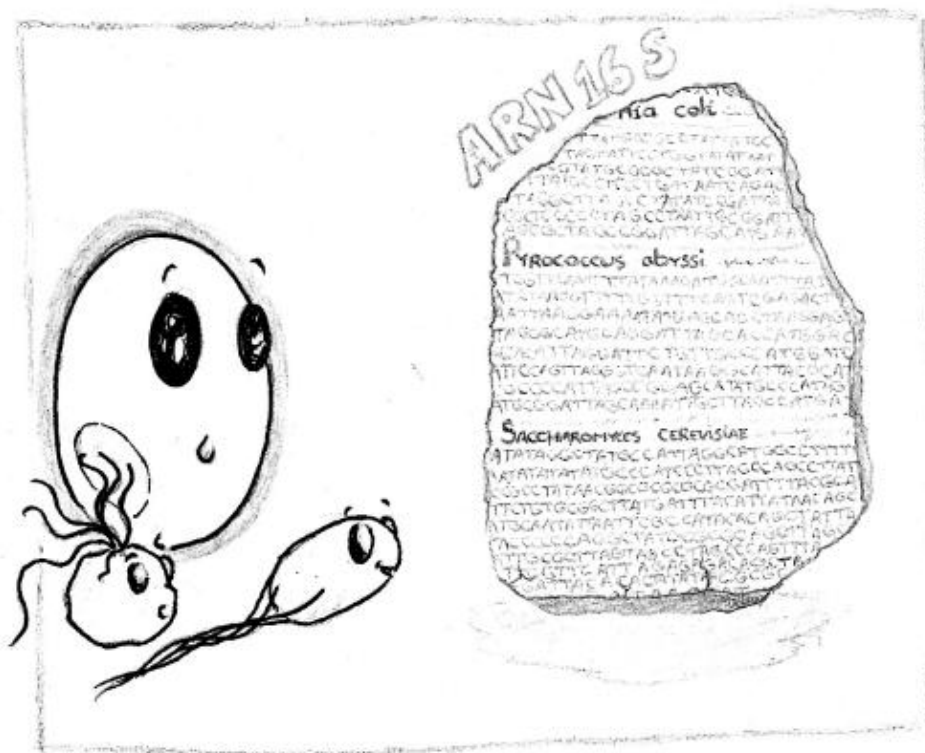
**FIGURE 3:** Les 5 règnes du vivant selon Whittaker (1969), adapté. Les Monères correspondent aux procaryotes, organismes unicellulaires de petite taille sans noyau apparent. Les *Protista* correspondent aux organismes eucaryotes unicellulaires. Les groupes *Plantae*, *Animalia* et *Fungi* correspondent aux organismes eucaryotes pluricellulaires.

Empruntant les idées pionnières d'Edouard Chatton (1883-1947) dont les travaux étaient restés dans l'ombre en 1937, Stanier et Van Niel distinguèrent les cellules procaryotes (pro=*avant* karyon=*noyau*) des cellules eucaryotes (eu=*vrai*). Ces derniers se divisent par mitose et possèdent un noyau entouré d'une membrane, un cytosquelette, et des organites individualisés. A l'opposé, les cellules procaryotes sont plus petites et ne présentent pas ces structures. Les Procaryotes furent ainsi définis de manière largement négative. Ils affirmèrent que ce groupe était monophylétique, bien que d'après eux, une classification phylogénétique de ces organismes resterait une chimère impossible et constituerait une « perte de temps ». La dichotomie Eucaryote/Procaryote du monde vivant fut alors rapidement ancrée parmi les dogmes de la biologie moderne, et constitue aujourd'hui les fondements de l'enseignement de cette discipline.

Ainsi, à la fin des années 1960, le règne des Monères, introduit par Haeckel, fut ressuscité. Représentant l'ensemble des Procaryotes par un groupe monophylétique, il fut placé comme l'un des cinq règnes du monde vivant, au même rang que les Protistes (Eucaryotes unicellulaires), des Animaux, des Plantes, des Champignons (Eucaryotes pluricellulaire) (R.H. WHITTAKER, 1969) (FIGURE 3)

Ce découpage du monde vivant avait l'avantage de fournir un scénario simple, élégant et gradualiste sur l'histoire évolutive de la vie. Les Monères ou Bactéries, avec leur petites cellules sans noyau, seraient des organismes primitifs, apparus les premiers (Procaryotes = avant le noyau) desquels auraient dérivés les Protistes eucaryotes unicellulaires, puis les trois grands règnes pluricellulaires. Le concept binaire *cellule primitive procaryote Vs cellule complexe eucaryote* constituait la trame sous jacente de cette histoire.

C'est à cette époque qu'une découverte fondamentale va faire éclater ces concepts, en introduisant un tout nouveau groupe d'êtres vivants, et en imposant une vision non plus dichotomique mais tripartite du monde vivant...



DESSIN ADDITIONNEL :

L'ARN 16/18S, Pierre de Rosette de la phylogénie des organismes



## **B - Jamais deux sans trois - Découverte d'un troisième domaine du vivant**

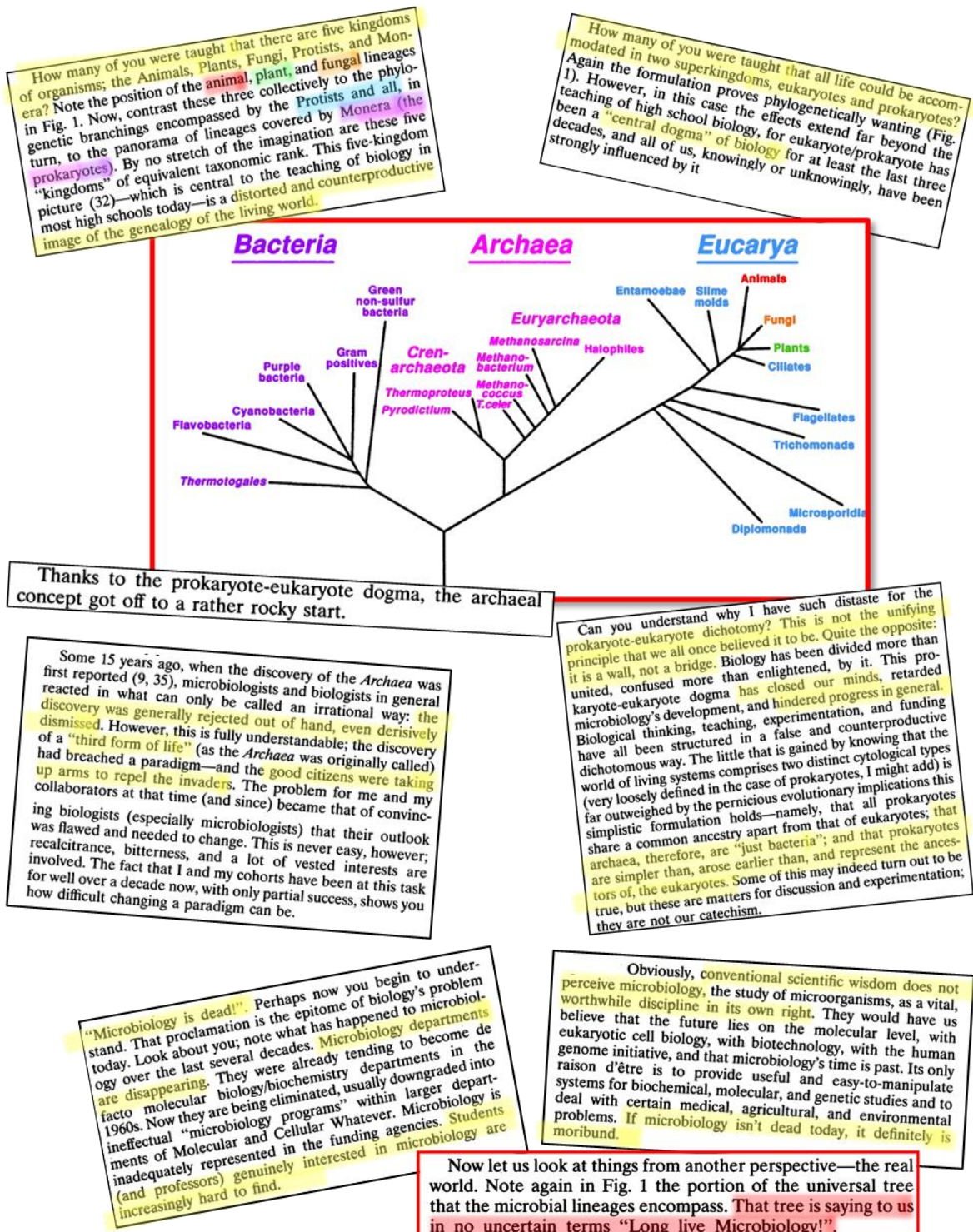
En 1977, un article d'à peine deux pages et demi, publié dans le journal PNAS par Carl Woese (1928-2012), et son post-doctorant George Fox, va jeter un pavé dans la mare bien ordonnée de la classification dichotomique du monde vivant (WOESE & FOX, 1977). Basée sur les récents développements des techniques de décodage de l'information contenue dans les acides nucléiques et les protéines (ZUCKERLAND ET PAULING, 1965), l'idée de construire une phylogénie universelle du vivant, basée sur la comparaison d'un seul et même caractère partagé au niveau moléculaire, devenait possible. Carl Woese, physicien de formation qui avait participé quelques années plus tôt à l'aventure du décodage du code génétique, et qui avait réalisé à cette occasion l'intérêt évolutif de la biologie moléculaire, se lança dans cette entreprise.

Il utilisa pour la première fois comme molécule informative l'ARN 16S/18S de la petite sous-unité du ribosome, présente chez tous les organismes cellulaires sans exception en grande quantité et facilement isolable. Avant que l'on ne développe des méthodes de séquençage à grande échelle, Carl Woese fit appel à des méthodes enzymatiques pour extraire l'information contenue dans cette molécule. Il établit, pour chaque ARN16S issu des différents organismes modèles, un catalogue d'oligonucléotides en faisant agir une enzyme clivant l'acide nucléique après chaque guanine. La comparaison de ces catalogues lui permit d'établir un coefficient de similarité ( $S_{ab}$ ) entre deux organismes donnés, reflétant leur proximité évolutive. Ces travaux révélèrent alors, à la grande surprise, que les bactéries méthanogènes, des Procaryotes connus pour leur métabolisme original (anaérobie et réduction du  $\text{CO}_2$  en méthane), étaient, au niveau de leur ARN16S, aussi différentes des autres bactéries que des Eucaryotes.

Woese et Fox proposèrent donc leur idée révolutionnaire : le monde vivant serait divisé en trois grands règnes primaires (urkingdom). Les « Eubactéries » (correspondant à la majorité des Procaryotes connues à l'époque), les « Eu/urkcaryotes » (correspondant aux Eucaryotes, sans prendre en compte leurs organites) et les « Archeobactéries », ce dernier règne n'étant constitué à cette époque que de ces méthanogènes. Cette proposition était également soutenue par certaines particularités biochimiques uniques de ces derniers.

MICROBIOLOGICAL REVIEWS, Mar. 1994, p. 1-9  
**There Must Be a Prokaryote Somewhere:  
 Microbiology's Search for Itself†**

CARL R. WOESE\*



**FIGURE 4 :** Extraits de l'article de C. Woese, « There must be a prokaryote somewhere », publié en 1994. L'auteur y combat la vision dichotomique du monde vivant et défend l'existence des Archées et l'importance des micro-organismes, en commentant l'arbre du vivant à trois branches établi par des méthodes de phylogénie moléculaire.

Le terme « archéobactérie » (anciennes bactéries) provient justement du métabolisme atypique des méthanogènes : il était supposé refléter les conditions régnant il y a 3 à 4 milliard d'années sur Terre. Ce nom fut remplacé au profit d'Archée dans les années 1990 (WOESE ET KANDLER ET AL, 1990), afin d'éviter de semer toute confusion avec le domaine des Bactéries. Depuis cette découverte, grâce à des décennies d'échantillonnages, de chasse aux Archées aux quatre coins du monde, d'analyse des génomes, de caractérisation biochimique, ce groupe s'est révélé être aussi riche en diversité que les deux autres groupes « classiques » que sont les Bactéries et les Eucaryotes. Il est donc bel et bien établi aujourd'hui que le monde vivant comporte trois grandes lignées cellulaires distinctes, et que le terme Procaryote cache en réalité deux lignées à part entière. Les données issues de la recherche sur les Archées ont fourni des avancées majeures dans de multiples champs de la biologie, et en fournissent encore. Outre le fait d'avoir refoulé une vision classique dichotomique, la découverte de C.Woese remet en question nos certitudes sur l'histoire des organismes cellulaires que nous avons esquissée. Il pose la question brûlante de l'origine de ces trois lignées cellulaires et de leur lien évolutif. Il raviva aussi les cendres d'une classification phylogénique des Procaryotes, depuis longtemps abandonnée (Van Stiel qualifiant de perte de temps), et marqua d'une pierre blanche l'ère de la phylogénie moléculaire.

En dépit de son importance, l'idée d'une division tripartite du monde vivant a eu du mal à être intégrée aux paradigmes de la biologie moderne. La dichotomie Eucaryote/Procaryote a ses défenseurs virulents (MAYR, 1998), certains esprits se refusant à voir les Archées autrement que comme de simple bactéries, certes atypiques mais ne justifiant pas l'établissement d'un domaine à part entière, et C.Woese n'a cessé de se battre pour faire entendre sa conception du vivant (WOESE, 1994) (FIGURE 4). Encore aujourd'hui le terme Procaryote, employé si fréquemment, masque la réalité pourtant concrète de l'existence des Archées, et continue d'instaurer une vision négative des micro-organismes. A terme, on peut espérer que le concept des trois domaines du vivant finisse par s'imposer, comme certaines notions fondamentales dépassant le clivage binaire ont fini par s'imposer dans les concepts humains : la sainte trinité, les trois couleurs primaires, les trois lois de la robotique, le pouvoir des trois...



## C - Rencontre du troisième type - Aperçu général des Archées :

Afin de comprendre l'impact de l'existence d'un troisième domaine du vivant dans les concepts fondamentaux de la biologie, il est nécessaire d'avoir un aperçu de ce que sont les Archées et en quoi elles constituent véritablement un domaine du vivant bien distinct.

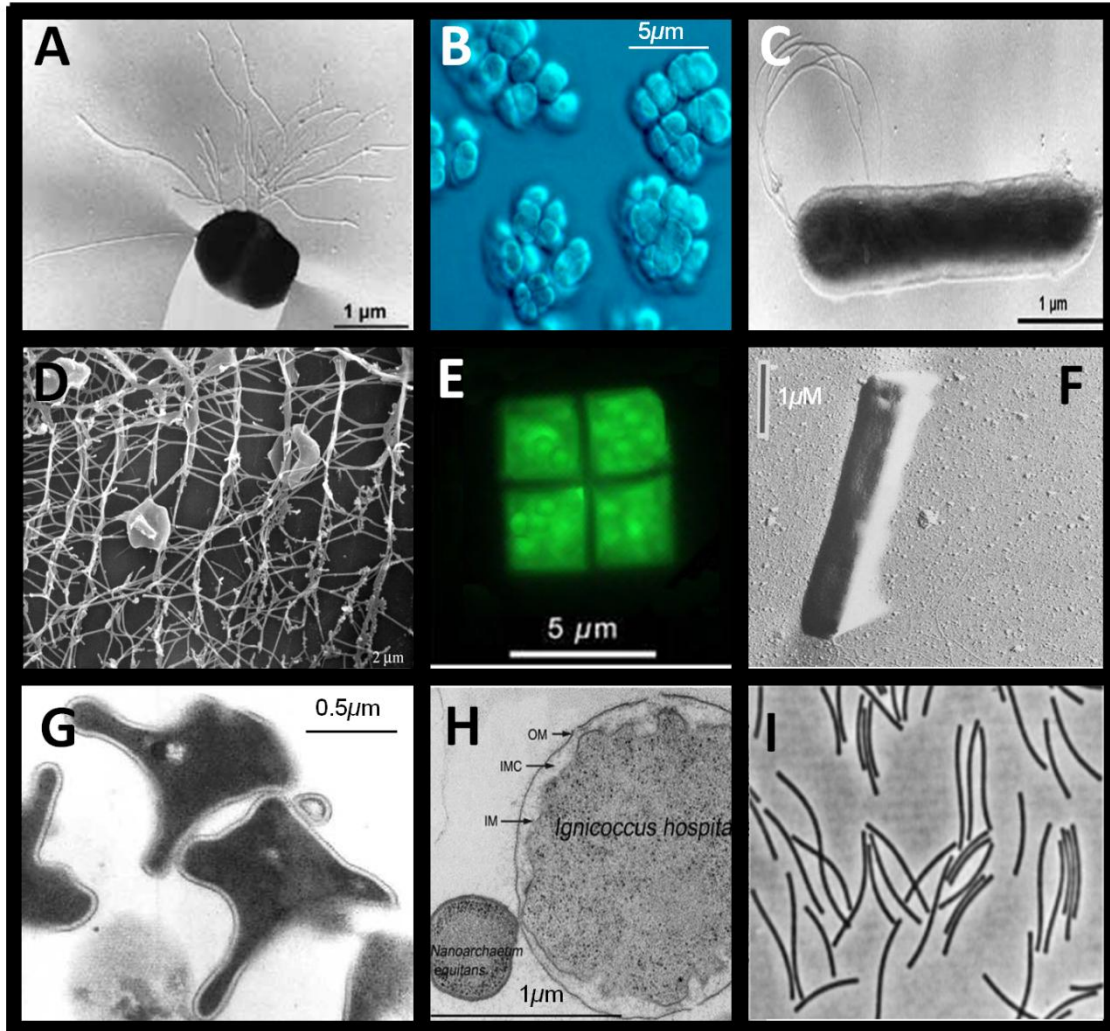
### 1 - Certains l'aiment chaud - Diversité écologique des Archées

Les premières Archées identifiées furent donc historiquement les Méthanogènes. Ces organismes étaient déjà connus pour leur biotope et leur métabolisme original : une absence stricte d'oxygène et une production de méthane à partir de composés carbonés.

Rapidement, d'autres organismes vinrent grossir le rang des Archées. En effet, on avait déjà identifié d'autres « bactéries » atypiques, ne possédant pas de peptidoglycane, et résistantes aux antibiotiques classiques comme la pénicilline (WOESE ET AL., 1978). L'analyse de leur catalogue d'oligonucléotides confirma leur proximité avec les Méthanogènes. Ainsi, les Halophiles extrêmes, retrouvés dans les milieux d'une extrême salinité comme la Mer morte ; et les organismes qualifiés de thermo-acidophiles, ayant un métabolisme basé sur le soufre isolés dans des sources chaudes à proximité des zones volcaniques (*Sulfolobus solfataricus*, isolée dans la commune de Solfatara, à côté de Naples et *Thermoplasma acidophilum*, isolée dans un terril acide d'une mine de charbon) furent également identifiés comme étant des Archées par Carl Woese (MAGRUM ET WOESE, 1978 pour *Halobacterium halobium*, WOESE ET AL., 1980 pour *Thermoplasma*, WOESE ET AL., 1984 pour *Sulfolobus*).

Il semblait donc que la caractéristique majeure de ce nouveau domaine du vivant était un certain amour pour les environnements extrêmes. La chasse aux organismes thermophiles (au-delà de 60°C) et hyperthermophiles (au-delà de 80°C), lancée par Karl Stetter et Wolfram Zillig (1925-2005), vint renforcer cette idée : à mesure qu'on l'on battait des records de thermophilie, on réalisait que ce domaine était bel et bien le royaume des Archées. Ainsi, *Pyrococcus furiosus*, isolé en 1986 dans les sédiments brûlants des bords d'une plage de l'île de Vulcano en Italie (FIALA & STETTER, 1986), et son cousin *Pyrococcus abyssi*, isolé en 1992 dans le conduit d'une cheminée hydrothermale océanique, à 3500 mètres de profondeur (ERAUSO ET AL., 1992), se plaisent à croître à 90°C, aux côtés du recordman *Pyrolobus fumarii* qui croît jusqu'à 113°C (BLOCHL ET AL., 1997).





**FIGURE 5 : Aperçu de la diversité morphologique des Archées.**

A. *Pyrococcus furiosus* (K. Stetter),

B. *Methanosarcina acetivorans* (Everly Conway de Macario),

C. *Aquifex aeolicus* (Stetter)

D. *Pyrodictium abyssi* (Rieger),

E. *Haloquadratum walsbyi* (Mike Dyll-Smith)

F. *Pyrobaculum aerophilum* (Völk),

G. *Sulfolobus acidocaldarius* (Grogan),

H. *Ignicoccus hospitalis* et *Nanoarchaeum equitans* (Eichinger),

I. *Methanospirillum hungatei* (Preslott).

Note: Une Bactérie s'est glissée (délibérément) parmi ces photographies...

Au bilan, la quasi totalité des organismes hyperthermophiles (c'est-à-dire se développant au-delà de 80°C) sont des Archées, et seuls quelques genres de Bactéries sont adaptés à ces environnements infernaux (les genres *Aquifex*, *Thermotoga* et *Thermus*).

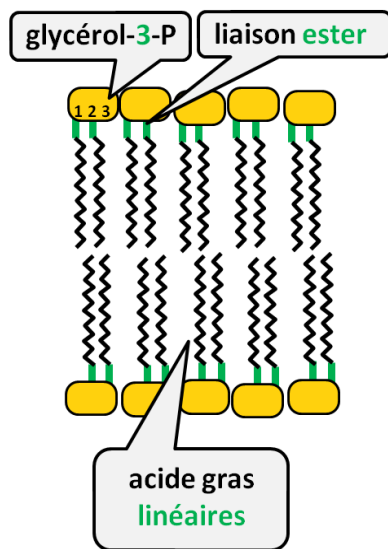
Par ces découvertes et bien d'autres, le caractère extrêmophile colle encore à la peau, ou plutôt à la membrane des Archées. Pourtant, des Archées mésophiles et aérobies furent également isolées, il est vrai bien plus tard (*Cenarchaeum symbiosium* isolé en 1996, et *Nitrosopumilus maritimus*, isolé en 2005). Finalement, les résultats issus de la métagénomique environnementale, par séquençage massif de l'ARN16S, ont révélé une facette méconnue de ce domaine : les Archées sont présentes dans tous les biotopes, et constituent probablement des acteurs non négligeables des écosystèmes). Par exemple, on estime qu'au niveau des océans entre 100 et 5000 mètres de profondeur, la densité de cellules d'Archées planctoniques serait de  $10^5$  cellules/mL, ce qui représenterait 20% de la biomasse microbienne de l'océan (DELONG & PACE, 2001). La compréhension de leur rôle écologique est donc aujourd'hui l'un des enjeux majeurs.

Toutes ces Archées, cultivables ou identifiées par métagénomique, sont classées en plusieurs grandes familles, selon trois phylums principaux : les Crénarchées, les Euryarchées et les Thaumarchées (BROCHIER-ARMANET ET AL. 2008).

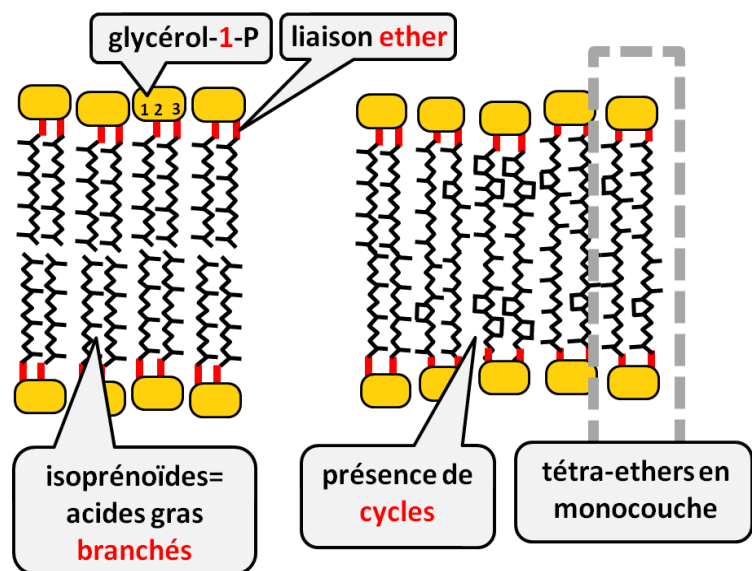
## 2 - Une apparence trompeuse - les Archées, sosies des Bactéries ?

D'un point de vue strictement cytologique, il est évident que les Archées se distinguent difficilement des Bactéries (FIGURE 5). C'est d'ailleurs l'une des raisons de leur découverte tardive. Ce sont des organismes unicellulaires, d'une taille moyenne de l'ordre du micromètre, présentant des morphologies variées : en forme de bâtonnet, de coques, carrés... qui se divisent par scission binaire, et qui forment des biofilms. Leur cytoplasme n'est pas compartimenté, leur chromosome est circulaire, d'une taille comprise entre 0,5Mb (*Nanoarchaeum equitans*) à 5,8Mb (*Methanosarcina mazei*) avec une organisation comparable à celle des Bactéries, et de nombreuses Archées possèdent des plasmides extra chromosomiques. Elles présentent une diversité métabolique assez variée : chimio/organotrophe, auto/hétérotrophe, aérobie ou anaérobie.

## Membrane des Bactéries et des Eucaryotes



## Membranes des Archées



**FIGURE 6 :** La membrane des Archées, une spécificité unique dans le monde vivant

Représentation schématique simplifiée de la membrane lipidique classique des Eucaryotes et Bactéries (à gauche), comparée à celle des Archées. La représentation au centre est une membrane composée du lipide le plus répandu chez les Archées, basé sur le glycérol-diéther appelé archaeol. La représentation de droite illustre des lipides tétraéther issus de la condensation de deux lipides diéther, et pouvant présenter des cycles carbonés.

Les principales différences chimiques entre les deux types de membranes sont indiquées. La partie supérieure correspond aux différences stéréochimiques au niveau du glycérol, la partie inférieure illustre des différences au niveau des chaînes d'acides gras (d'après Matsumi *et al*, 2011)

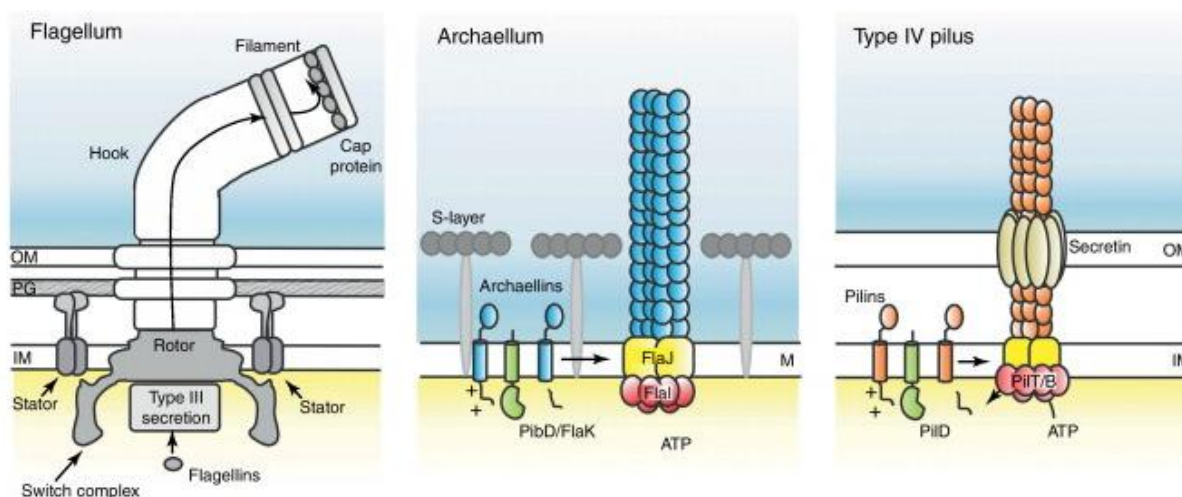


Hormis leur extrêmophilie, qui n'est en fait pas nécessairement la norme, il est donc difficile au niveau conceptuel (pour certains microbiologistes) de prendre appui uniquement sur l'ARN 16S pour justifier d'un domaine à part : en quoi les Archées sont-elles si différentes des Bactéries auxquelles on tient tant à les distinguer ? C'est en réalité au niveau biochimique et génomique que se trouve la réponse.

### 3 - Des cellules pas comme les autres - Spécificité des Archées

Avant même leur identification en tant que groupe distinct, les premières Archées isolées : Méthanogènes, Thermoacidophiles et Halophiles, que l'on prenait pour des Bactéries, étaient connues pour certaines caractéristiques biochimiques singulières : une enveloppe ayant une composition particulière, des co-enzymes uniques, une insensibilité surprenante à certains antibiotiques bactériens... Depuis des décennies de caractérisations, de très nombreuses nouvelles spécificités archéennes ont été décrites ; dont quelques unes sont particulièrement marquantes et ne laissent aucun doute quant à la singularité des Archées.

L'une des caractéristiques majeures qui distingue les Archées de tous les autres organismes vivants se trouve au niveau de leur membrane. Celle-ci est composée de lipides totalement uniques, tant par leur composition que par leur stéréochimie (**FIGURE 6**) (**BOUCHER ET AL. 2004**). Chez les Bactéries comme chez les Eucaryotes, les lipides membranaires sont constitués d'un groupement polaire lié à un phosphate, lui-même lié à une molécule de glycérol, au niveau du carbone 1. Les deux autres carbones (2 et 3) du glycérol sont liés par liaison ester à des chaînes d'acides gras linéaires. Chez toutes les Archées, le phosphate est lié au glycérol au niveau du carbone 3, et les carbones 1 et 2 sont liés par des liaisons éther (et non pas ester) à des chaînes d'alcool aliphatiques complexes, dérivés de l'isoprényl (on dit qu'ils sont branchés car ils comportent des groupements méthyles latéraux) et pouvant comporter des cycles à 5 ou 6 carbones. Enfin, certains lipides sont des tétraéthers, issus de la condensation de deux lipides par l'extrémité de leur chaîne aliphatique. La quasi-totalité des livres de biologie enseignent encore aujourd'hui un modèle universel de bicouche lipidique formé de phospholipides classiques : en réalité, parler de la membrane des « Procaryotes » n'a plus de sens au vu de celle, totalement atypique, des Archées.



**FIGURE 7 :** Comparaison entre le flagelle bactérien, le « flagelle » archéen et le pili de type IV des Bactéries

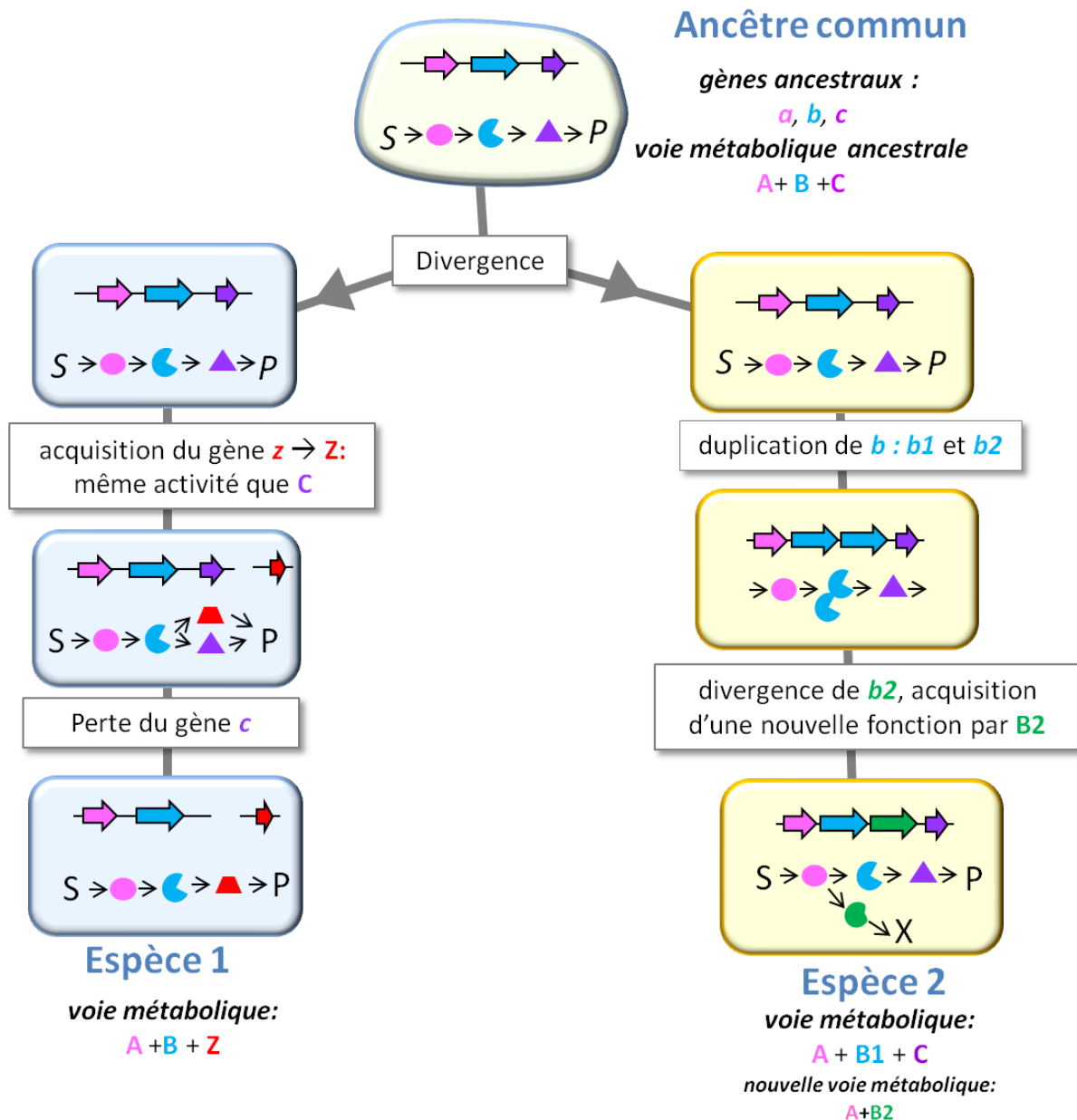
L'illustration met en évidence les différences de composition et de structure entre les deux types de flagelle, et montre la similarité entre le flagelle des Archées (ou archaeellum) et le pili de type IV des Bactéries. Le schéma permet également d'avoir un aperçu de la S-Layer des Archées, formé de protéine et glycoprotéine semi-cristalline. D'après S Albers, 2011.

Chez la plupart des Bactéries, la membrane plasmique est entourée d'une enveloppe rigide constituée de peptidoglycane, assemblage de monomères de sucres et de peptides. Les Archées, en revanche, présentent une très grande diversité d'enveloppes, chimiquement et structurellement très différente celles des Bactéries. L'enveloppe la plus répandue chez les Archées est appelé *S-layer*, composée d'un seul type de protéine ou de glycoprotéine organisée de manière semi-cristalline et directement associée à la membrane plasmique (ALBERS ET AL. 2011).

Autre exemple : certaines Archées se déplacent dans leur milieu, tout comme les Bactéries, à l'aide d'une structure rotative appelée flagelle (ou « flagellum »). Or, le flagelle des Archées n'a rien à voir avec celui des Bactéries, tant par sa composition, par son mode de formation et son fonctionnement, et il est en réalité proche des pili de type IV. Il a ainsi été proposé récemment d'appeler cette structure l'« archaellum » afin de bien distinguer cette structure de son analogue bactérien (JARRELL & ALBERS, 2012) (FIGURE 7).

Enfin, dernier exemple du caractère unique de ce domaine : certaines Archées méthanogènes ont introduit dans leur code génétique un acide aminé totalement inédit : la pyrrolysine, en réassignant le codon stop ambre UAG. Cet acide aminé est utilisé dans les protéines impliquées dans la méthanogénèse (qui est également un métabolisme exclusivement archéen) (ATKINS ET AL, 2002).

**Au niveau biochimique, les Archées représentent bien un type cellulaire clairement unique. Toutes ces particularités traduisent l'existence d'enzymes spécifiques qui n'ont pas d'équivalents chez les Bactéries et les Eucaryotes. Les Archées ont donc leur propre histoire évolutive, distincte des autres organismes. Cependant, certains caractères et processus sont partagés à la fois par les Archées, les Bactéries et les Eucaryotes : l'ARN 16S, qui a mené à la découverte des Archées, en est l'exemple même. La comparaison de ces caractères communs peut donc fournir des renseignements sur les liens évolutifs des trois lignées, et sur l'histoire évolutive du vivant d'une manière générale.**



**FIGURE 8:** Illustration des notions d'orthologie et de paralogie.

A partir d'un ancêtre hypothétique contenant trois gènes ( $a, b, c$ ) codant des protéines ( $A, B, C$ ) impliquées dans une même voie métabolique, deux chemins évolutifs sont proposés aboutissant à deux espèces différentes. La branche n°1 illustre la notion de déplacement de gène: le gène  $c$  est remplacé par le gène  $z$ , codant une protéine  $Z$  ayant la même fonction que la protéine  $C$  dans la voie métabolique. La branche n°2 illustre la notion de duplication et de paralogie. Le gène  $b$  subit un événement de duplication, conduisant à l'apparition de deux copies identiques  $b1$  et  $b2$ . Le gène  $b2$  diverge par perte de pression de sélection, permettant l'apparition d'une protéine  $B2$  possédant une nouvelle fonction.

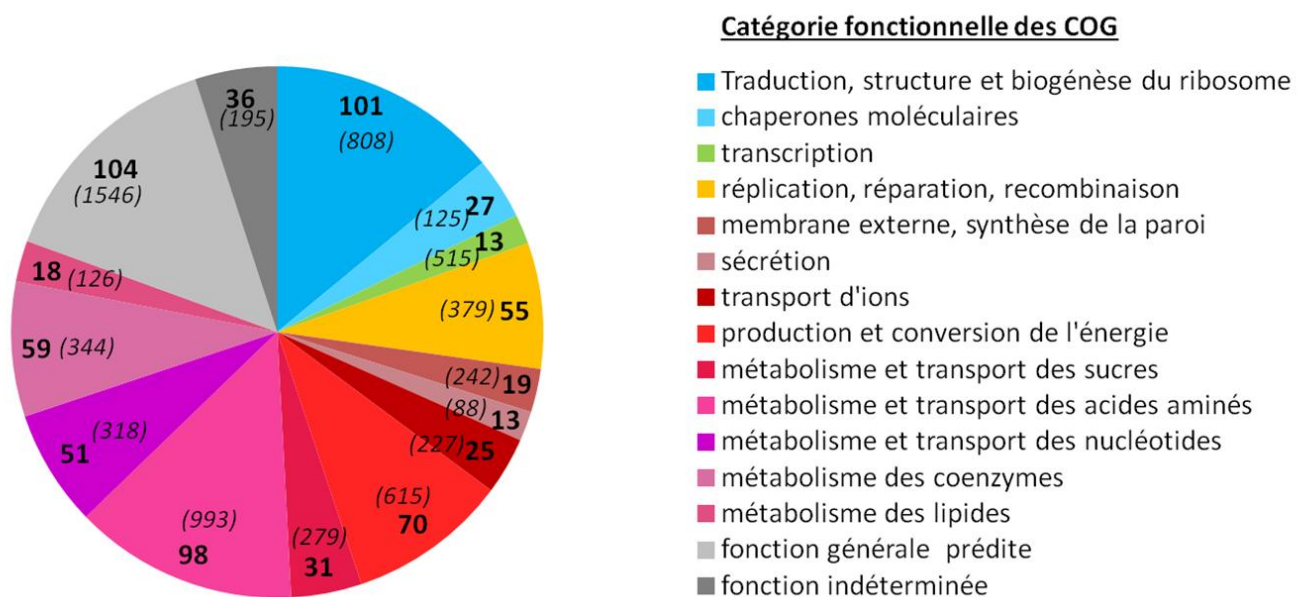
- Les séquences  $a$  (ou  $b$ ) de l'espèce 1 et 2 sont des **orthologues**
- Les séquences  $b1$  et  $b2$  de l'espèce 2 sont des **paralogues**
- $z$  et  $c$  ne sont pas homologues: les protéines  $Z$  et  $C$  sont des **analogues fonctionnels**

## 4 - Comment comparer Archées, Eucaryotes et Bactéries : notions de génomique comparative

Historiquement, ce fut donc la séquence de l'ARN16S, molécule universelle, qui fut utilisée par Carl Woese pour comparer les êtres vivants ; fournissant ainsi la première étude globale de phylogénie moléculaire. Cette molécule présente de nombreux avantages, et les techniques d'isolation, d'identification et de séquençage de cet ARN16S ont été rapidement développées, fournissant année après année des dizaines, centaines, milliers de nouvelles séquences. L'arbre initial à trois branches s'étoffa donc progressivement de nombreux rameaux secondaires (WOESE ET AL, 1987), pour devenir une référence absolue (bien que controversée sous plusieurs aspects) en terme de phylogénie.

Cependant, l'utilisation de l'ARN16S présente certaines limites en termes d'information que l'on peut en tirer. Grâce aux progrès du séquençage des génomes qui connut un essor fulgurant dans les années 1990, on disposa alors d'autres marqueurs moléculaires utilisables pour la comparaison des organismes, ce qui marqua la naissance de la génomique comparative. Avant de revenir aux surprises que réservent encore les Archées au niveau de leurs gènes et de leur comparaison avec les deux autres domaines du vivant, il est nécessaire de définir les notions fondamentales de cette approche.

La génomique comparative repose sur la comparaison des séquences présentes chez deux ou plusieurs génomes, en se basant sur leur degré de ressemblance. Concrètement, la comparaison des séquences d'un gène ou d'une protéine permet d'établir un pourcentage d'identité et de similarité. On estime que des séquences qui partagent un pourcentage suffisant de similarité témoignent d'une origine commune, c'est-à-dire qu'elles dérivent d'une même séquence ancestrale : on dit qu'elles sont homologues. Deux principales notions découlent de celle de l'homologie : l'orthologie et la paralogie. (FIGURE 8) (KOONIN, 2005, revue complète sur les notions d'orthologie et de paralogie) Deux gènes sont dits orthologues lorsqu'ils sont présents dans deux espèces différentes qui ont divergé par spéciation à partir d'un même ancêtre commun. Deux gènes sont dits paralogues lorsqu'ils sont issus d'une duplication d'un gène ancestral, au sein d'un même génome.



**FIGURE 9 : Répertoire original des Groupes de Gènes Orthologues (COG) établi en 1997**

A partir du contenu de 7 génomes entièrement séquencés : 5 Bactéries (*Escherichia coli*, *Haemophilus influenzae*, *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Synechocystis sp.*), 1 Archée (*Methanocaldococcus jannaschii*) et 1 Eucaryote (*Saccharomyces cerevisiae*), les gènes ont été regroupés par ensemble de gènes orthologues : les COG. Un COG est constitué lorsque trois lignées phylogénétiques au moins présentent un gène orthologue. Les paralogues (duplication au sein d'une même lignée) de ce gène, lorsqu'ils existent, sont inclus dans le même COG. Les différents COG obtenus (720, contenant 6814 protéines, soit 37% des gènes de l'ensemble des 7 génomes), ont été ensuite regroupés en grandes catégories fonctionnelle, selon la fonction prédites ou vérifiée des protéines qu'ils contiennent. Cette distribution est représentée ici sous la forme d'un diagramme circulaire. Le nombre en gras indique le nombre de COG pour chaque catégorie, le nombre de protéines correspondant est indiqué entre parenthèses (d'après Tatusov *et al*, 1997).

Généralement, les gènes orthologues codent pour des protéines accomplissant des fonctions similaires d'une espèce à l'autre, alors que des gènes paralogues ont souvent des fonctions proches mais différentes, l'un des deux gènes acquérant de nouvelles caractéristiques par perte de pression de sélection issue de la présence simultanée de deux copies. Ainsi, l'identification de séquences orthologues peut permettre de déduire et prédire la fonction d'un gène chez un organisme, à condition d'avoir déjà caractérisé le rôle de son homologue chez une autre espèce. Elle permet aussi de retracer l'histoire évolutive des gènes, par différentes méthodes de phylogénie moléculaire.

De nombreuses méthodes et bases de données ont ainsi été développées à partir des années 1990 pour classer, organiser et analyser les informations issues du séquençage des génomes. Parmi ces outils, nous allons présenter les COGs, pour *Cluster of Orthologous Groups (ou Genes)*, car leur utilisation a été à l'origine des certaines découvertes concernant les travaux présentés dans cette thèse. Cette base de données a été établie en 1997, par comparaison de séquences des gènes (basé sur le calcul d'indice de similarité, le Best Hit=BeT) afin de dresser une liste des gènes conservés dans les génomes d'organismes phylogénétiquement éloignés (7 génomes complets étaient alors disponibles, 1 Archée, 1 Eucaryote et 5 Bactéries) (FIGURE 9). Chaque COG est défini comme l'ensemble des « gènes orthologues individuels ou groupes orthologues de gènes paralogues, détectés au minimum dans 3 lignées évolutivement éloignées » (TATSUTOV ET AL, 1997). Chaque COG est ainsi supposé avoir évolué à partir d'un unique gène ancestral par événements de spéciation et de duplications. On suppose alors que les membres d'un même COG remplissent des fonctions similaires ou proches en raison de leur relation d'homologie, et les COGs sont classés en grands ensembles fonctionnels.

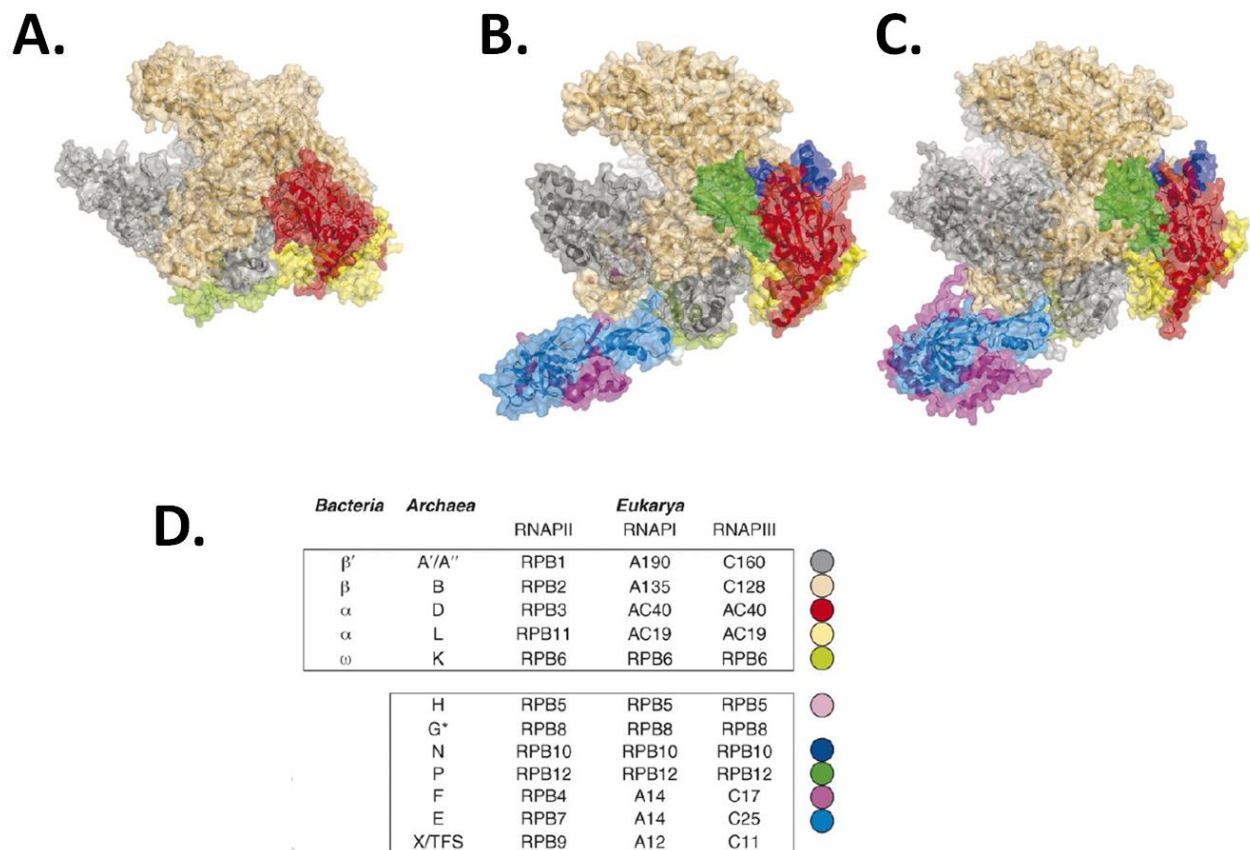
A mesure que le nombre de génomes séquencés augmentait, la base de données des COGs s'est étoffée au fil des ans (720 COGs en 1997 pour 7 génomes, 2091 en 2000 pour 21 génomes, et 4873 COGs pour 66 génomes en 2003), et à force de se complexifier et de s'enrichir, des bases plus spécialisées ont été développées : les COG spécifiques aux Eucaryotes (KOG) (TATUSOV ET AL, 2003), ceux spécifiques aux Archées (arCOG) (actualisé en 2012, WOLF ET AL). Ces banques de données ont été un outil puissant pour l'annotation des nouveaux génomes, comme par exemple celui des Archées hyperthermophiles *Pyrococcus*





*abyssi*, *Aeropyrum pernix* (NATALE ET AL., 2000), *Candidatus Korarchaeum cryptofilum* (ELKINS ET AL., 2008). Les COG ont également été employés pour étudier la distribution de certains gènes et proposer des scénarii évolutifs. Par exemple, il est apparu que la reverse-gyrase est la seule protéine commune à tous les organismes hyperthermophiles, ce qui suggère que cette enzyme joue un rôle fondamental pour l'adaptation aux très hautes températures (FORTERRE, 2002). Enfin, l'utilisation des COG a été utilisée pour déterminer un « Top 10 » des protéines devant faire l'objet d'une étude expérimentale prioritaire, dont nous reparlerons par la suite (KOONIN & GALPERIN, 2004, 2010).

Cependant, la comparaison des séquences par des logiciels automatiques et l'utilisation des COG n'est pas un oracle absolu. Certains gènes peuvent ne pas être identifiés comme orthologues si leurs séquences respectives ont trop divergé. D'autres gènes peuvent avoir fusionné ou au contraire s'être scindés, ce qui nécessite une vérification manuelle. De plus, la règle orthologie = conservation de la fonction n'est pas une règle absolue, et la fonction des paralogues est difficile à prédire. Enfin, les cas de perte de gène, les transferts horizontaux de gènes entre organismes peuvent brouiller les pistes sur la relation des gènes. Seule la main (et l'intuition) de l'expérimentateur peut confirmer les prédictions issues des logiciels de génomique comparative, et interpréter de manière correcte les nombreuses données qui en résultent.



**FIGURE 10:** Comparaison structurale des ARN polymérases (RNAP) issues des trois domaines du vivant.

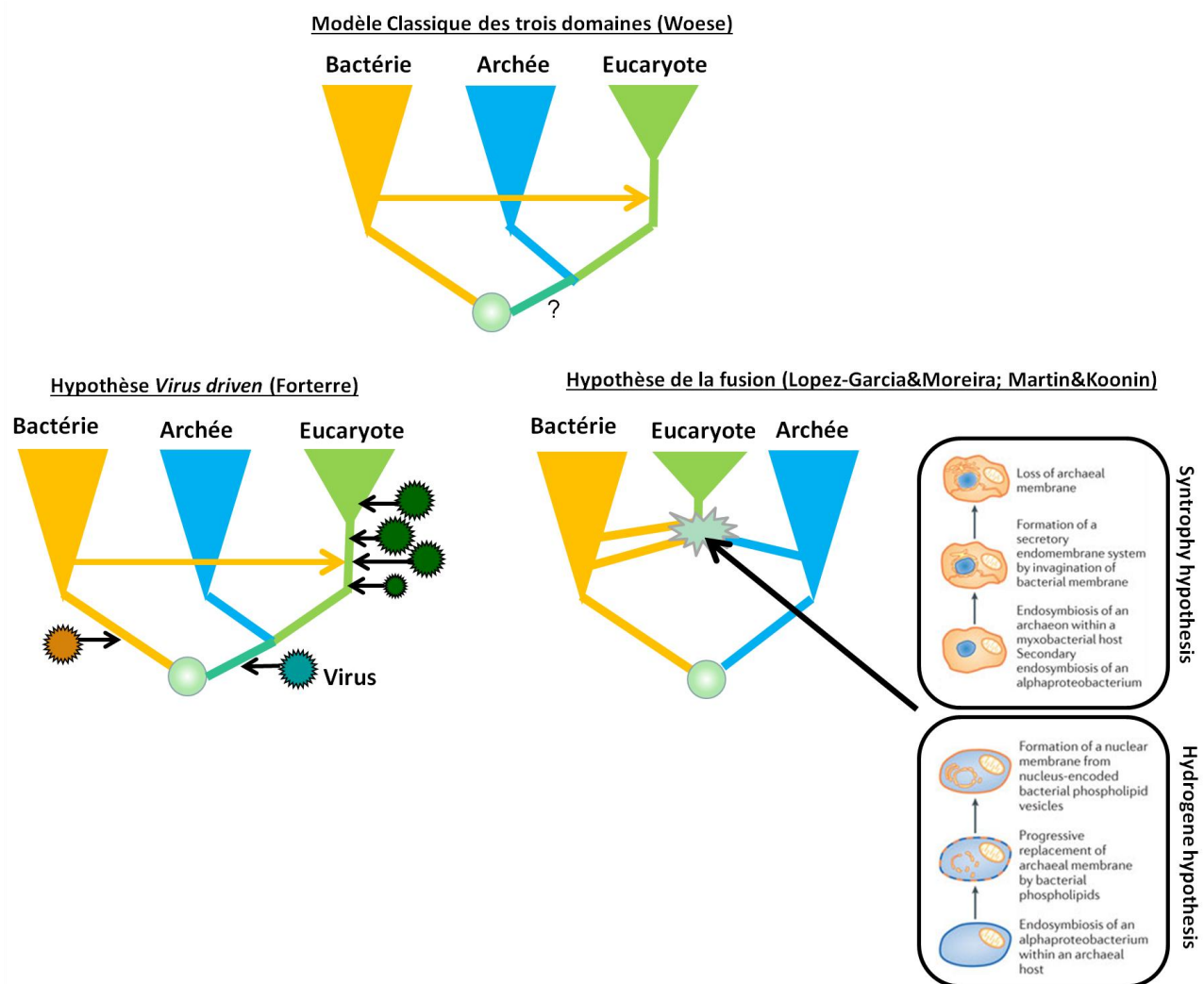
A. RNAP de la Bactérie *Thermus aquaticus*. B. RNAP de l'Archée *Sulfolobus solfataricus*. C. L'ARN polymérase II de l'Eucaryote *Saccharomyces cerevisiae*. D. Composition des ARN polymérases des trois domaines du vivant. Les sous-unités sont classées en RNAP (colonnes), selon leur degré d'homologie de séquence et de structure (lignes). Le cadre supérieur indique les sous-unités universelles, le cadre du bas indique les sous-unités communes aux Archées et aux Eucaryotes. Les cercles colorés correspondent aux couleurs employées en A, B et C. D'après F.Werner, 2008

## 5 - L'habit ne fait pas le moine - L'étonnante proximité entre Archées et Eucaryotes au niveau moléculaire

Plusieurs études biochimiques, réalisées avant cette avalanche de génome, avaient déjà soulignées une ressemblance frappante entre certains processus chez les Archées et les Eucaryotes, au niveau de leurs protéines (ZILLIG ET AL. 1978, au sujet de l'ARN polymérase de *Sulfolobus*). L'avènement de la génomique comparative, qui révéla de manière très exhaustive la distribution des gènes conservés entre les différents domaines du vivant, vint soutenir ces observations et permit de généraliser cette idée : les Archées et les Eucaryotes partagent au niveau moléculaire de nombreux points communs. En voici certains exemples.

Les Bactéries sont sensibles à la rifampicine. Cet antibiotique inhibe leur ARN polymérase, enzyme multiprotéique responsable de la transcription. Les Eucaryotes, en revanche, y sont insensibles, car leurs ARN polymérases diffèrent fortement de celles des Bactéries, et possèdent notamment un nombre plus important de sous-unités. De manière surprenante, les Archées sont également insensibles à la rifampicine (PRANGISHVILLI ET AL, 1982). La caractérisation de l'ARN polymérase d'Archée permet d'expliquer cette observation : elle présente presque autant de sous-unités que les polymérases Eucaryotes, et ces sous-unités sont orthologues (FIGURE 10). De même, le mécanisme de transcription est également semblable, bien que simplifié chez les Archées, avec l'utilisation d'une séquence TATA où se lie une TBP (*Tata Binding Protein*), orthologue à la TBP des Eucaryotes (THOMM ET AL., 1996)

De même, les Bactéries sont sensibles à l'action du chloramphénicol et de la kanamycine, qui bloquent la traduction en perturbant le fonctionnement de leur ribosome. Les Eucaryotes y sont insensibles, en raison de différences structurales au niveau de leur ribosome. Les Archées sont également insensibles à l'action de ces antibiotiques. La caractérisation du ribosome d'Archée montre ici aussi une très forte ressemblance avec celui des Eucaryotes. En plus d'un cœur commun de protéines ribosomales universelles, les Archées et les Eucaryotes partagent 33 protéines homologues que ne possèdent pas les Bactéries (LONDEI, 2007, « ARCHAEA : MOLECULAR AND CELLULAR BIOLOGY, CHAP. « TRANSLATION »).



**FIGURE 11:** Exemples de scénarii simplifiés expliquant les liens et l'évolution des trois domaines du vivant.

L'ancêtre commun, LUCA ou Cenancestor, est placé à la racine de l'arbre (rond vert). B= Bactéries, A= Archées, E= Eucaryotes. V= virus. La flèche orange symbolise l'évènement d'endosymbiose qui a donné naissance à la mitochondrie.

Dans le modèle classique, les trois lignées ont divergés à partir de l'ancêtre, les Archées et les Eucaryotes partageant un ancêtre commun plus récent qu'avec les Bactéries. La position exacte de la racine est sujette à controverse (?)

Dans l'hypothèse de la fusion, il n'y a que deux domaines primaires (Procaryotes: Archées et Bactéries): les Eucaryotes proviennent d'une fusion entre une archée et une ou plusieurs bactéries, selon les hypothèses (encarts). Le modèle *virus-driven* propose que les virus ont joué un rôle actif et déterminant dans l'évolution des trois lignées, notamment dans l'apparition des Eucaryotes (par exemple dans l'apparition de l'ADN, dans l'apparition du noyau). D'après Woese, Lopez-Garcia Moreira, Forterre

D'autres scénarios alternatifs ont été proposés.

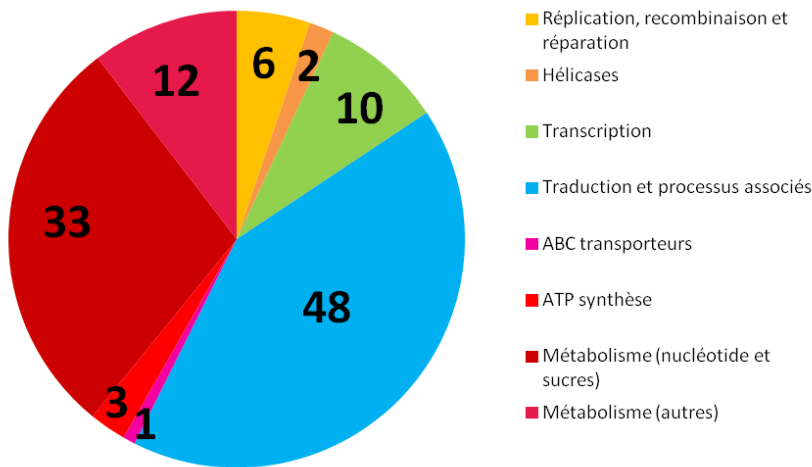
Cette proximité au niveau moléculaire entre Archées et Eucaryotes est retrouvée dans de très nombreux autres processus cellulaires: réparation et réplication de l'ADN, production d'ATP, dynamique cellulaire (cytosquelette et dynamique membranaire). Dans tous ces processus, les protéines d'Archées sont des orthologues proches des protéines Eucaryotes (R.CAVICCHIOLI, 2007, « ARCHAEA : MOLECULAR AND CELLULAR BIOLOGY »). L'étude de ces protéines chez les Archées constitue ainsi un excellent modèle pour comprendre les mécanismes moléculaires à l'œuvre chez les Eucaryotes. Le complexe KEOPS, que l'on retrouve chez les Archées et les Eucaryotes et qui sera caractérisé dans cette thèse, en sera un exemple.

## D - La génomique comparative, une fenêtre sur le passé :

Cette proximité entre les Archées et les Eucaryotes révèle une histoire évolutive commune entre ces deux domaines du vivant. L'apparition des trois grandes lignées cellulaires modernes apparaît donc plus complexe qu'on ne le pensait, et de nombreux scénarios évolutifs ont été proposés pour expliquer les liens qui unissent Archées, Eucaryotes et Bactéries (FIGURE 11). L'étude comparative des processus universellement à l'œuvre chez les organismes modernes permet ainsi de remonter leur état ancestral, et d'esquisser un portrait-robot du dernier ancêtre commun universel : LUCA. Plusieurs questions se posent alors : quelles étaient les caractéristiques de LUCA par rapport aux cellules modernes ? Quels sont les processus cellulaires déjà en place à son époque, quels étaient ses composants, et comment ont-ils évolué dans les trois grandes lignées ?

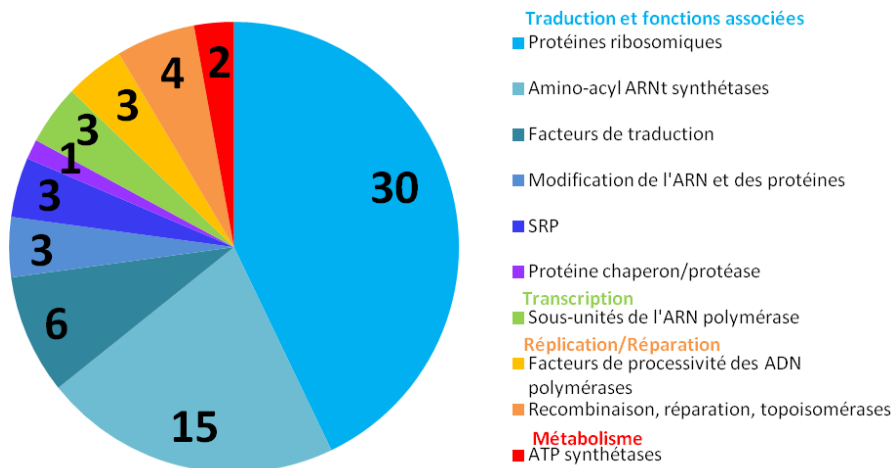
Pour répondre à ces questions, trois approches complémentaires sont à considérer: l'étude des gènes universels, l'étude des gènes essentiels, et l'étude des gènes présents chez les organismes dits « minimalistes ». Ces différentes approches et les questions évolutives auxquelles elles sont rattachées seront abordées au cours de ces travaux de thèse.

A.



Famille Kae1/YgjD (COG0533) absente car non détectée chez *M. jannaschii*  
 Famille YrdC/Sua5 (COG009) : présente dans traduction

B. Gènes universellement conservés



Famille Kae1= présente, annotée chaperon/protéase  
 Famille YrdC= absente car non détectée chez *Helicobacter*

### FIGURE 12 : Distribution des gènes conservés entre différentes espèces.

Les gènes ont été regroupés en catégorie fonctionnelles indiqués à côtés des diagrammes. Les couleurs utilisées correspondant aux couleurs des COG de la figure 8.

A= gènes conservés entre la bactérie *E. coli*, l'Archée *M. jannaschii* et l'Eucaryote *S. cerevisiae*, d'après Delaye 2005.

B= gènes universellement conservés déterminés par la comparaison des COG dans les génomes entièrement séquencés en 2003. D'après Forterre 2005, Koonin 2003

Des informations (en rose et en violet) concernant les deux familles de protéines universelles dont il est question dans ce manuscrit ont été ajoutées en bas de chaque diagramme.

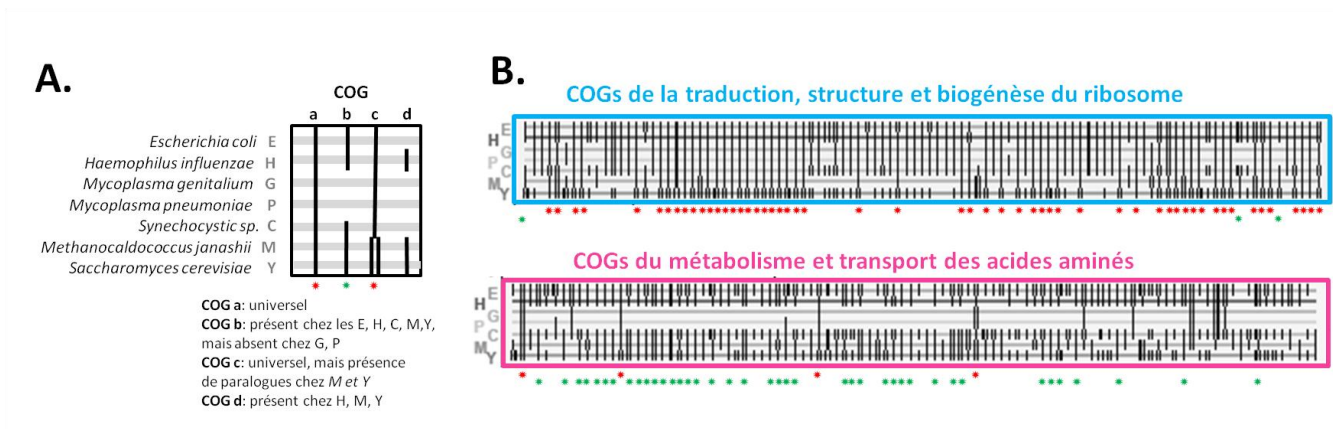
## 1 - Le cœur des génomes - Les gènes universels :

Afin d'identifier les processus conservés et de remonter à leur état ancestral, une première approche consiste en l'identification des « génomes-cœur » (*core genome* en anglais). Par génomique comparative, on peut établir la liste des protéines conservées entre deux ou plusieurs organismes, et établir *ad fine* une liste de protéines universelles.

Cette approche fut utilisée pour la première fois en 1996, dès que les deux premiers génomes de bactéries furent disponibles, celui d'*Haemophilus influenzae* et *Mycoplasma genitalium*, (MUSHEGIAN&KONIN, 1996), puis plusieurs études tentèrent d'identifier les gènes universellement conservés pour déterminer ceux qui devaient déjà exister chez l'ancêtre commun (HARRIS ET AL., 2003, KONIN ET AL, 2003, PACE ET AL, 2003). La FIGURE 12-A illustre le résultat obtenu par comparaison de trois organismes modèles issus de chacun des domaines du vivant, et la FIGURE 12-B illustre l'un des résultats obtenu par comparaison de tous les génomes séquencés en 2003. Deux paradoxes sautent alors aux yeux : d'une part, sur les milliers de COG identifiés, seuls 70 sont universels (selon les études, ce chiffre varie entre 60 et 80). D'autre part, alors que la majorité des COG identifiés concernent le métabolisme, seulement deux sont retrouvés dans le panel de COG universels. LUCA était-il donc un organisme à 70 gènes, n'ayant quasiment aucun métabolisme?

En réalité, l'analyse des COG, et particulièrement des COG universels, a ses limites. En fonction des génomes comparés, et surtout de leur nombre, la taille du *génome-cœur* varie fortement. D'une manière générale, plus le nombre de génome que l'on compare est important, plus le panel de gènes conservés est réduit, en raison des pertes, des transferts horizontaux et des déplacements de gènes. Ces phénomènes touchent particulièrement les gènes impliqués dans le métabolisme, qui sont fréquemment échangés ou perdus contrairement aux gènes impliqués dans la machinerie de traduction (FIGURE 13).

Le résultat obtenu par l'étude des gènes universels est donc par nature minimaliste, et ne reflète pas l'état complet de l'ancêtre hypothétique. Certaines conclusions peuvent toutefois être tirées (retour sur la FIGURE 12) : par exemple, on s'aperçoit que les protéines de réplication de l'ADN, processus pourtant universel, sont presque absentes de cette liste. Cela vient du fait que la machinerie bactérienne n'a aucun lien d'homologie avec la machinerie



**FIGURE 13 : Illustration du biais induit par la comparaison des COG entre 7 organismes**

Les deux catégories fonctionnelles majoritaires: la traduction (en bleu) et le métabolisme des acides aminés (en rose) qui présentent environ le même nombre de COGs (101 contre 98) ont été choisies comme exemple. Les lignes correspondent aux 7 génomes de références (5 bactéries, E, H, G, P, C; 1 Archée M, 1 Eucaryote S), les colonnes aux différents COGs.

La présence d'un COG dans chacun de ces 7 génomes est représentée par une barre verticale noire. Une barre verticale unique indique qu'un seul gène représentant le COG en question est présent. Une double barre verticale indique qu'au moins deux paralogues sont présents au sein du génome. Une étoile rouge symbolise un COG retrouvé chez les 7 génomes (qualifié d'universel), une étoile verte indique les COG retrouvés chez tous les génomes, en ne prenant plus en compte les Mycoplasmes.

A) Par exemple, les 7 génomes présentent chacun un unique représentant du COG a: ce COG est qualifié d'universel. En revanche, seule E, H, C, M et Y possèdent un représentant du COG b. Pour le COG c, chaque organisme possède un seul représentant de ce COG, sauf M et Y, qui présentent plusieurs paralogues. Le COG d n'est présent que chez H, M et Y.

B) On constate que la plupart des COGs de la traduction sont universels. En revanche, pour les COG du métabolisme des acides aminés, on constate que la prise en compte des Mycoplasmes (G et P), qui ne possèdent qu'une dizaine de COGs, induit un biais: la prise en compte de ces deux organismes réduit fortement le nombre de COG universels de cette catégorie, (d'après Tatusov, 1997)

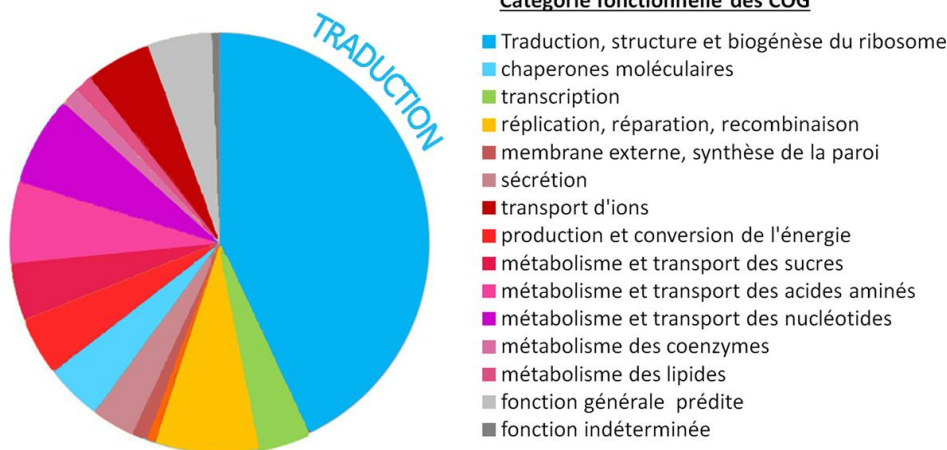


des Archées et Eucaryotes, et cette observation constitue un argument permettant de suggérer que l'ADN aurait été « inventé » deux fois indépendamment au cours de l'évolution et que LUCA aurait pu avoir un génome à ARN. (MUSHEGIAN&KONIN, 1996, FORTERRE, 2006). A l'opposé, on s'aperçoit que le nombre de gènes impliqués dans la traduction reste constant, quel que soit le nombre d'organismes considérés, et que ce nombre est assez important. Ainsi, la traduction est un processus universel hautement conservé, très ancien, et LUCA devait posséder un système traductionnel probablement déjà très sophistiqué et proche des systèmes modernes actuels (FORTERRE ET AL, 2005). L'un des enjeux de cette thèse concernera cette question.

## 2 - Jamais sans mon gène ! - Les gènes essentiels :

La recherche des gènes hautement conservés a donc ses limites, car de nombreux événements au cours de l'évolution ont effacé en partie l'héritage issu de LUCA; et il faut faire appel à des approches complémentaires pour continuer à dresser son portrait-robot. L'une d'elle consiste à identifier les gènes essentiels dans les cellules modernes, au moyen d'études mutationnelles. Un gène est qualifié d'essentiel lorsque sa délétion entraîne la mort de l'organisme en question, et la liste des gènes essentiels chez différents organismes permet de définir les réseaux métaboliques qui constituent le fondement de la vie cellulaire. Une banque de données de gènes essentiels confirmés et prédits a été ainsi mise en place: la *Database of essential gene* (DEG <http://tubic.tju.edu.cn/deg/>) et *database of predicted essential gene* (pDEG) (ZHANG ET AL., 2004). La FIGURE 14-A illustre le résultat obtenu par comparaison des gènes essentiels et conservés dans le génome de 18 espèces de Mycoplasmes. D'après l'analyse de ces données, il apparaît d'une manière générale que la plupart des très nombreux gènes liés à la traduction sont des gènes essentiels chez tous les organismes, ce qui est en accord avec les observations précédentes (ZHANG ET AL., 2009) la traduction est un processus complexe dont les différents éléments ont dû se mettre en place sous leur forme moderne très tôt dans l'évolution, pour devenir indispensables à la cellule, et sont résistants aux pertes de gènes.

A.



B.

Species	Number of protein-coding genes	Number of genes analysed	Number of essential genes	Number of essential genes extrapolated to complete genome	Gene-inactivation method	References
<i>M. genitalium</i> / <i>M. pneumoniae</i>	480	All (random mutagenesis)	351 without insert	265–380 (55–79%)	Transposon-insertion mutagenesis	31
<i>B. subtilis</i> <sup>†</sup>	4,118	3,613	192	271 (6.6%)	Plasmid-insertion mutagenesis	33
<i>H. influenzae</i>	1,714	1,272	478	670 (38%)	Antisense-mediated gene inactivation	32
<i>E. coli</i>	4,275	3,746	620	708 (17%)	Transposon-insertion mutagenesis	15
<i>S. cerevisiae</i>	~6,000	5,916	1,105	1,124 (~19%)	Precise deletion by mitotic recombination	70
<i>C. elegans</i>	~20,000	16,757	929	1,080 (~5.4%)	Inactivation by RNA interference	71

\*All experiments assayed the survival of the respective organisms on rich media. <sup>†</sup>Combined with previously published data.

### FIGURE 14 : Etude des gènes essentiels

**A.** Distribution des 153 gènes constituant le «cœur» des gènes essentiels chez les Mycoplasmes, classés par COG fonctionnels. Ces 153 gènes correspondent à l'intersection des ensembles de gènes essentiels chez 18 génomes de Mycoplasmes (5880 prédits dans 16 génomes, 379 et 310 déterminés expérimentalement chez *M. genitalium* et *M. pulmonis*, respectivement). Les gènes impliqués dans la traduction (en bleu) sont majoritaires.

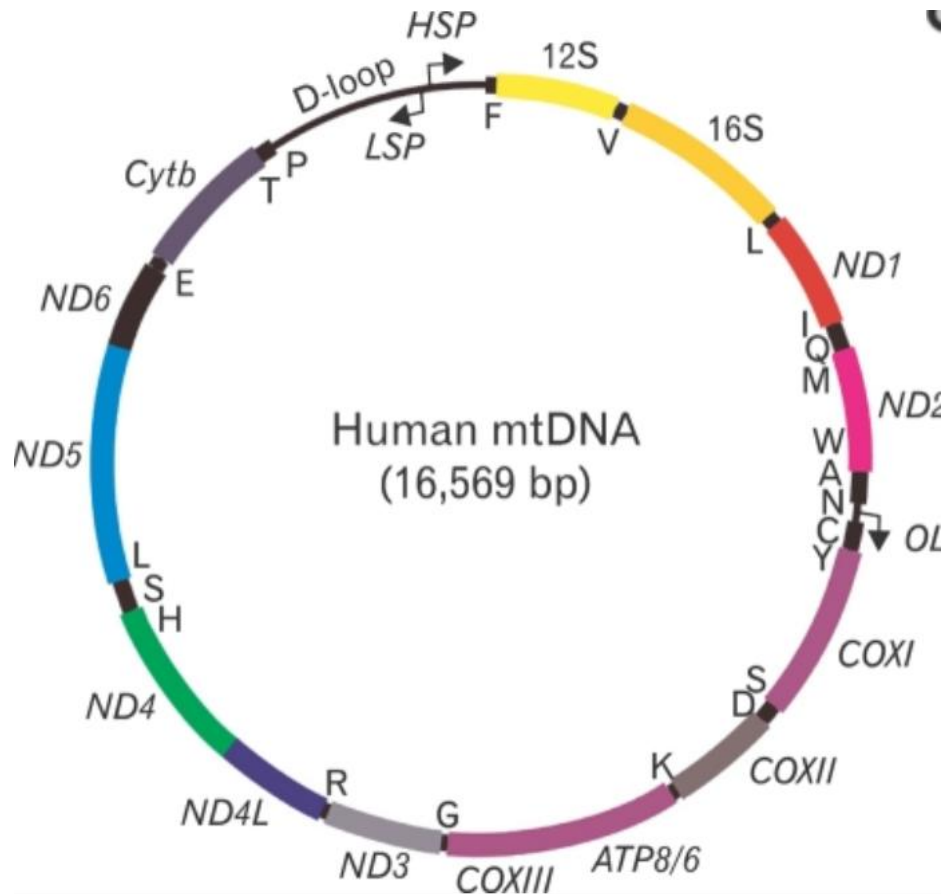
**B.** Tableau des gènes essentiels identifiés par inactivation chez les Bactéries et les Eucaryotes. d'après Lin et *al.*, 2011, et extraits de Database of predicted essential genes pDEG : <http://tubic.tju.edu.cn/pdeg//index.html>

On s'aperçoit pourtant avec surprise que le pourcentage de gènes qualifiés d'essentiels varie fortement d'un organisme à un autre (**FIGURE 14-B**) (17% chez *E. coli* à près de 80% chez *Mycoplasma*). Cela s'explique par plusieurs raisons. D'une part, elle dépend fortement des conditions de vie de l'organisme : un gène peut apparaître comme essentiel dans une condition donnée, et ne plus l'être dans une autre. D'autre part, les protéines formant des réseaux d'interactions complexes, deux gènes peuvent apparaître comme non essentiels lorsqu'ils sont délétés séparément, alors que leur délétion simultanée engendre un organisme non viable. De même, un gène orthologue peut être essentiel chez un organisme sans qu'il le soit chez un autre. Ainsi, le contexte génétique, l'espèce et l'environnement sont à prendre en considération dans la notion d'essentialité. Enfin, la technique employée est également critique pour qualifier un gène d'essentiel, car il faut s'assurer de l'inactivation totale du gène. Selon les études, certains gènes se sont ainsi vu qualifiés d'essentiels pour ne plus l'être quelques années plus tard, et le redevenir par la suite. Ces différents aspects seront abordés au sujet des protéines étudiées au cours de ces travaux, dont la question de l'essentialité est sujette à débats.

### 3 - Ce n'est pas la taille qui compte... Les génomes réduits :

Enfin, la dernière approche que l'on peut citer, et qui a été utilisée en combinaison des deux précédentes, est l'étude des organismes dits « minimalistes », dont le génome code pour un nombre réduit de protéines. Ces organismes peuvent nous renseigner directement sur le minimum requis pour construire et entretenir de façon viable une cellule.

*Pelagibacter ubique*, une  $\alpha$ -protéobactérie de l'ordre des Rickettsiales, est l'organisme libre qui détient le record du plus petit génome. Séquencé en 2005 (**GIOVANONNI ET AL., 2005**), il contient 1389 gènes identifiés, pour une taille de 1,3Mb. L'Archée *Thermoplasma acidophilum* le talonne, avec un génome de 1.56Mb, pour 1509 gènes (**RUEPP ET AL., 2000**). Viennent ensuite différents organismes qui ont tous en commun un mode de vie parasitaire (ou endosymbiotique) plus ou moins poussé. Les Mycoplasmes, qui parasitent des cellules de mammifères et sont un fléau pour les laboratoires de biologie cellulaire (**MARTIN ET AL., 2012**), sont utilisés comme modèle de système minimaliste dans de nombreuses études (**FIGURE 12-A**). Dans un milieu très riche, il est possible de faire croître, bien que difficilement, certains Mycoplasmes : ces organismes se situent donc à la limite d'une vie indépendante.



**FIGURE 15:** Représentation du génome mitochondrial humain, un exemple de génome très réduit.

Le génome contient deux ARN ribosomiques (12S et 16S, en jaune), 22 ARNt de transfert (indiqués par des lettres uniques) et 13 gènes essentiels qui codent les sous-unités du complexe enzymatique de phosphorylation oxydative. HSP et LSP= promoteurs de la région boucle D. OL = origine de la réplication du brin L. d'après Young *et al.*, 2010

Le génome de *Mycoplasma genitalium* contient 470 gènes pour 580kb (FRASER ET AL, 1995). L'Archée *Nanoarchaeum equitans* possède un génome de taille comparable, 552 gènes identifiés pour une taille de 491 kb (WATERS ET AL, 2003), mais cet organisme est un parasite/symbiote obligatoire de l'Archée *Ignicoccus hospitalis* et n'a jamais pu être isolé de son hôte. Le record absolu du génome le plus réduit est détenu par les bactéries qui vivent en endosymbiose dans des cellules d'insectes et se transmettent par héritage vertical. Le génome de *Carsonella ruddii*, qui code 159 protéines pour 160kb, est le cas le plus extrême, et cet organisme peut être envisagé comme un organite (NAKABACHI ET AL. 2006). Enfin, on peut mentionner les véritables organites que sont les mitochondries et les chloroplastes, lointains descendants de Bactéries. Totalement inféodés (emprisonnés ?) aux cellules Eucaryotes, ils ne sont pas considérés comme des organismes à part entière mais peuvent révéler des informations intéressantes, car ils possèdent encore un génome codant un certain nombre de protéines, et contiennent notamment des gènes d'ARNr et d'ARNt (FIGURE 15), ainsi qu'une machinerie de traduction fonctionnelle. Au cours de ces travaux, nous montreront l'intérêt de l'étude des processus chez ces modèles minimalistes, en caractérisant une protéine mitochondriale conservée. Nous reviendrons ainsi plus en détail dans une introduction dédiée à cet organite (Partie III)

En raison de leur mode de vie, ces différents organismes ont été soumis à un processus de réduction de leur génome au cours de l'évolution, se traduisant par une perte de gènes importante. Capable de se diviser, et présentant encore certaines caractéristiques cellulaires, ils sont cependant dépendants de leur hôte respectif pour de nombreux processus. Ainsi, l'étude de leurs gènes est nécessairement biaisée, et doit être abordée avec prudence, mais peut permettre d'identifier les processus ayant résistés à la perte de gène, et constituant par conséquent un cœur fondamental aux génomes, en synergie avec la recherche des gènes essentiels. Grâce à leur étude, on estime qu'une cellule minimale théorique devrait posséder entre 200 et 300 gènes (GIL ET AL. 2004), dont la majorité concerne encore une fois la traduction.



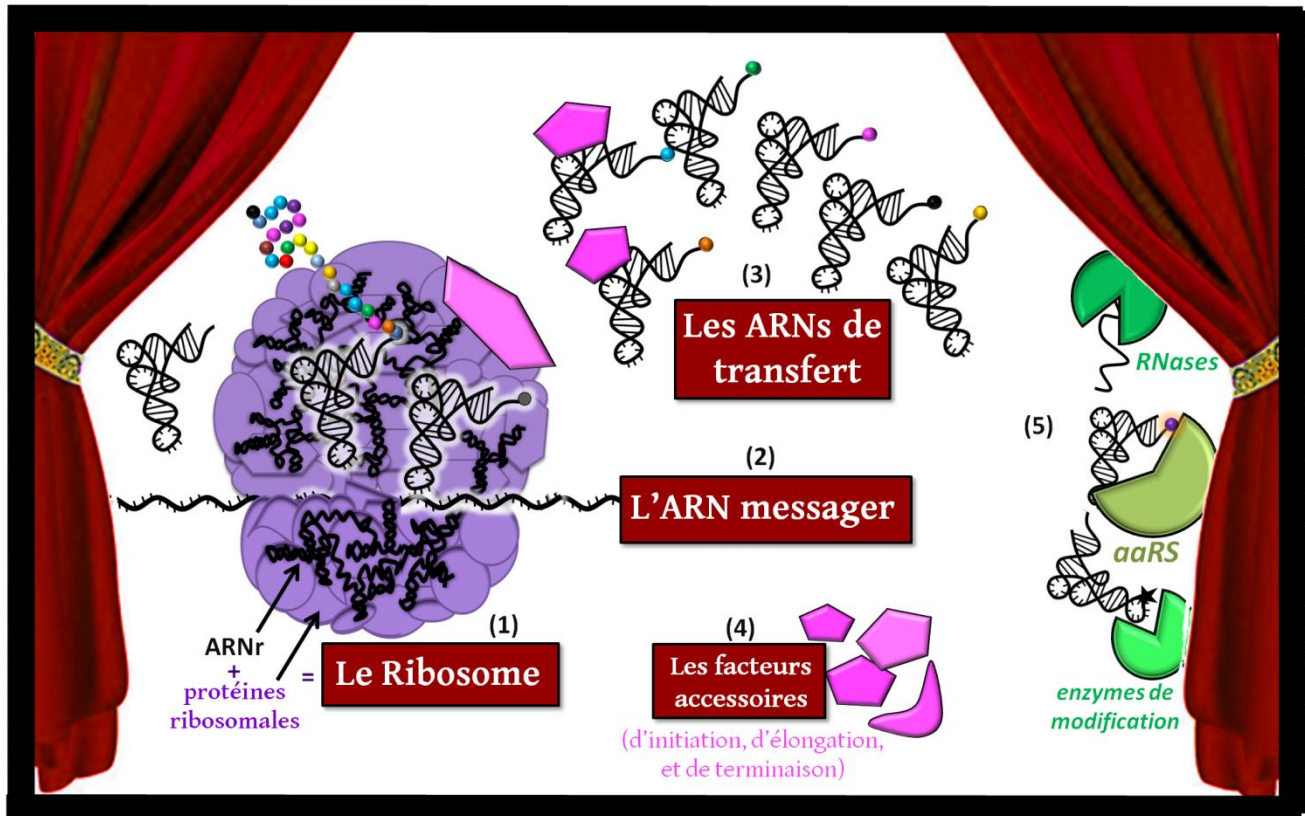
L'étude ayant mené à la découverte des Archées, premier pas de la phylogénie moléculaire, provoqua un bouleversement majeur dans deux disciplines de la biologie : la microbiologie, qui vit « apparaître » d'un coup un tout nouveau groupe d'organismes totalement singuliers, et la systématique, qui vit sa classification dichotomique du monde vivant se scinder en trois grandes lignées. Les Archées ont depuis enrichi nos connaissances dans de multiples domaines, de façons parfois inattendues. D'un point de vue évolutif, l'existence des Archées a également provoqué un ras de marée d'idées, d'hypothèses et de débats concernant l'ancêtre commun à tous les organismes vivants. La complexité moléculaire des Archées, et leur proximité avec les Eucaryotes, vint brouiller les pistes concernant l'idée que l'on pouvait se faire de cet ancêtre. Ainsi, on se tourna vers la génomique comparative afin de pouvoir identifier les processus cellulaires conservés, retracer leur histoire évolutive, et fournir un état probable ancestral de ces processus.

A l'issue des différentes études globales ayant identifié les gènes essentiels, universels et présents chez les organismes réduits, il apparaît clairement que la traduction est le processus central le plus conservé, qui résiste aux évolutions réductrices, aux pertes de gènes, aux transferts latéraux, et qui regroupe de manière surprenante un nombre impressionnant de gènes. La traduction est donc un processus ancien, qui devait probablement être déjà très complexe à l'époque de LUCA.

De plus, ces différents travaux montrèrent que seule une dizaine de protéines universelles restaient mystérieuses au début des années 2000: elles constituaient donc une priorité d'étude (GALPERIN&KONIN 2004, 2010).

L'objet de cette thèse concerne deux de ces protéines, qui semblent être impliquées dans un processus universel complexe lié à la traduction. Ainsi, avant d'aborder ce processus spécifique et les protéines qui feront l'objet de cette étude, une présentation générale de la traduction doit être faite.





**FIGURE 16: Le grand théâtre de la Traduction et ses principaux acteurs**

Le ribosome (1), composé d'ARNr et de protéines ribosomales, est la machinerie centrale de synthèse du polypeptide, où a lieu le décodage de l'information génétique, porté par l'ARN messager (2) sous forme de codons et où s'effectue la formation de la liaison peptidique. Les ARN de transfert (3) assurent ce décodage. Chaque ARNt porte un acide aminé spécifique et s'apparie avec l'ARNm au sein du ribosome selon les règles du code génétique qui assurent la correspondance entre codon et acide aminé. De nombreux facteurs protéiques (4) interviennent lors des étapes d'initiation, d'élongation et de terminaison.

Enfin, en coulisse s'affairent différentes classes d'enzymes (5) qui assurent la maturation des ARNm, ARNt, ARNr et protéines de ce système: différentes RNases, les aminoacylARNt synthétases, qui greffent sur chaque ARNt l'acide aminé qui lui correspond, et les enzymes de modifications, qui transforment chimiquement certains nucléotides des ARN, notamment des ARNt.

# Chapitre II- De la Traduction à t<sup>6</sup>A – Voyage au centre de l'ARNt

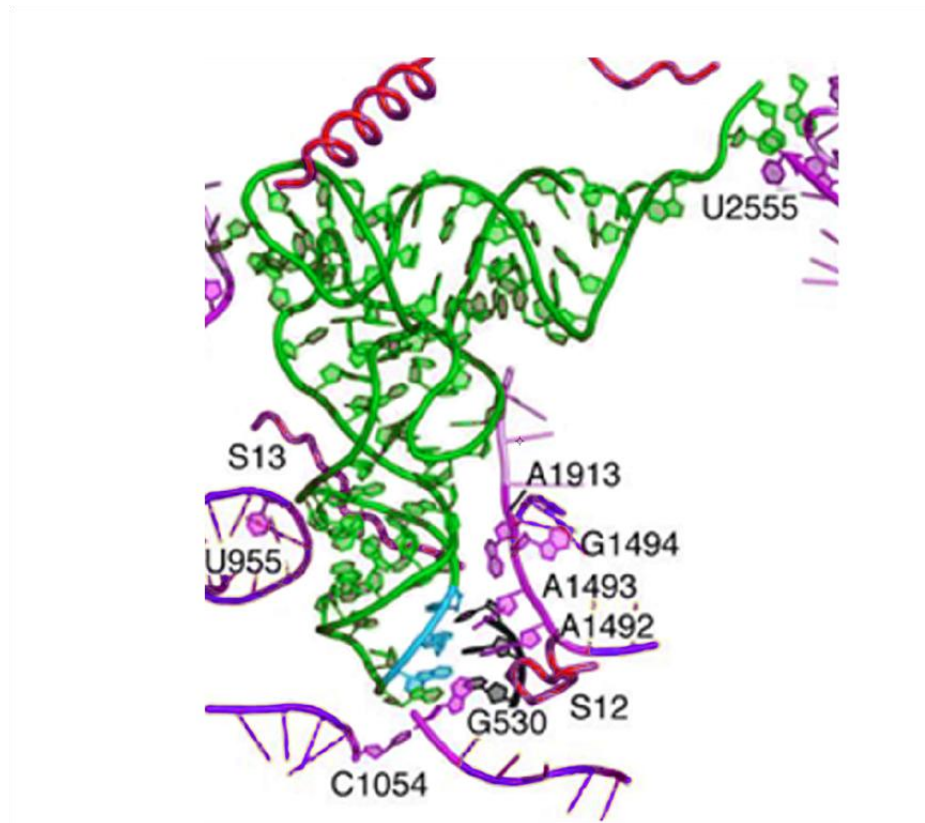
La traduction est l'étape clé du processus qui permet le passage de l'information stockée sous forme d'acide nucléique, le génotype, à des polypeptides matérialisant le phénotype. Décoder un acide nucléique implique d'une part l'utilisation d'un dictionnaire : le code génétique, mais aussi d'une machinerie spécifique capable de lire concrètement ce code. Ces deux aspects de la traduction apparaissent comme les éléments les plus conservés du monde vivant.

## A- Aperçu général de la traduction

### 1- Un ballet moléculaire universel

La traduction s'apparente à une pièce d'opéra (**FIGURE 16**) : des acteurs principaux, une troupe de ballet, une multitude d'acteurs secondaires et d'équipes de montage, et un programme en trois actes. Le scénario de cette pièce universelle suit des règles extrêmement précises et probablement établies avant la divergence des trois grandes lignées du vivant Archées, Eucaryotes et Bactéries, bien que chacune d'entre elles ait développé des variations spécifiques.

Les rôles principaux sont tenus par deux ténors : le ribosome (**1**), machinerie centrale où a lieu la synthèse des protéines, et composé à la fois d'ARN ribosomiques (ARNr) et de protéines ; et l'ARN messager (ARNm) (**2**), portant l'information génétique sous forme d'une succession de triplets de nucléotides appelés codons. Ils sont accompagnés par la troupe des différents ARN de transfert (ARNt) (**3**), et qui assurent la correspondance entre le langage nucléique et le langage protéique, en s'appariant à l'ARNm via le ribosome. Ces différents acteurs sont assistés de nombreux facteurs protéiques (**4**), pendant qu'en coulisse s'affairent une multitude d'enzymes de maturation (**5**). L'ensemble de la synthèse protéique s'effectue lors des trois Actes que sont l'Initiation, l'Elongation et la Terminaison.



**FIGURE 17 : Une danse à trois partenaires : représentation structurale d'un ARNt apparié avec un ARNm au sein du ribosome.**

Seul le codon de l'ARNm est visualisé en noir, apparié avec l'anticodon (en bleu) de l'ARNt (en vert). Les bases de l'ARNr (en violet et rose) impliquées dans le contrôle de l'appariement sont mises en évidence. Les bases A1493, A1492 et G530 de l'ARN16S interagissent avec la mini hélice du duplex codon/anticodon et permettent de contrôler l'appariement. (d'après *Cochella 2007*)

Lors de l'initiation, le ribosome engage la synthèse protéique au niveau du premier codon du cadre de lecture de l'ARNm, appelé codon initiateur, par l'utilisation d'un ARNt initiateur spécifique porteur d'une méthionine. Les mécanismes mis en jeu lors de cette étape, bien que reposant sur les mêmes principes, diffèrent chez les trois domaines du vivant (BENELLI ET AL., 2009). Les modalités d'assemblage du ribosome sur l'ARNm et de la reconnaissance du premier codon d'initiation font partie de ces variations, et sont l'objet de nombreuses études. Notamment, alors que le codon d'initiation employé chez les Eucaryotes est systématiquement un AUG, les Bactéries et les Archées peuvent initier leur traduction au niveau d'un AUG (majoritairement), mais aussi d'un GUG ou d'un UUG, en utilisant le même ARNt initiateur (FILES ET AL. 1975, KHUDYAKOV ET AL., 1985, YANG ET AL., 1999). Cette différence entre Eucaryotes et Archées/Bactéries trouve son origine probablement au niveau de cet ARNt particulier, mais n'est pour le moment pas totalement élucidée.

Deuxième acte : l'élongation, qui est l'étape la plus conservée au niveau mécanistique et dont la chorégraphie minutieuse est digne des plus grandes scènes. Une fois la première méthionine mise en place, les acides aminés sont assemblés en un polypeptide par le ribosome grâce à la lecture des codons successifs. Les ARNt dit élongateurs entrent tour à tour au sein du ribosome. Chaque ARNt porte en extrémité 3' un acide aminé spécifique, et possède un triplet de nucléotide complémentaire appelé anticodon, qui s'apparie au codon de l'ARNm engagé dans le ribosome. Une fois cet appariement effectué, l'acide aminé porté par l'ARNt est transféré au polypeptide en formation lors d'une réaction de transpeptidylatation, catalysée par certaines bases de l'ARNr de la grande sous-unité : le ribosome est en réalité un ribozyme, probablement la relique d'un ancien monde à ARN (FOX, 2010).

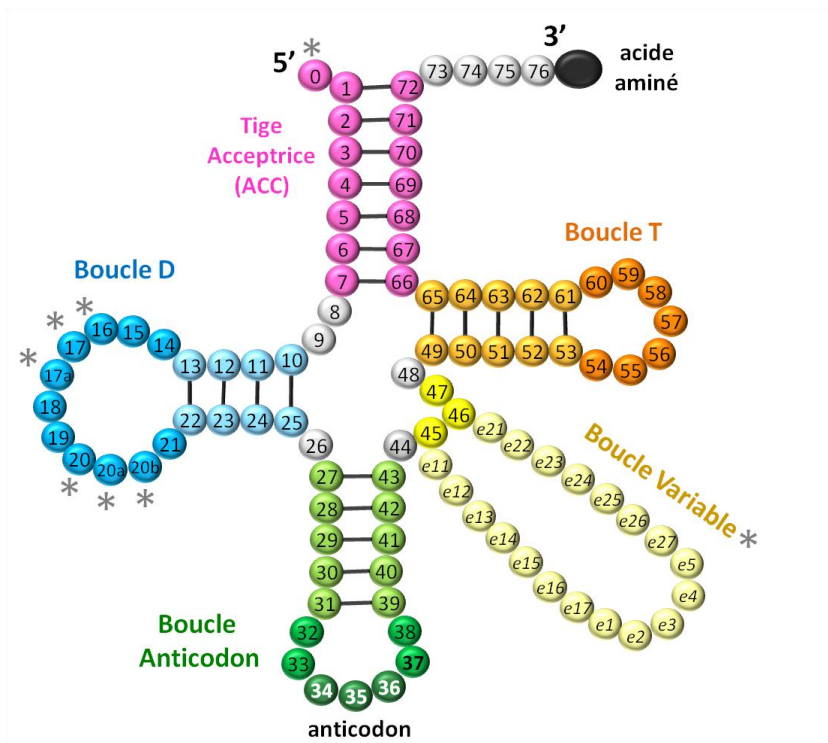
La précision de cette étape est indispensable pour obtenir une traduction sans erreur (chez *S. cerevisiae*, le taux d'erreur est d'environ  $10^{-4}$  erreur par codon, et chez *E. coli*  $10^{-3}$  to  $10^{-4}$ ) (KRAMER ET AL., 2010, KRAMER ET AL., 2007), et les ARNt entrant au sein du ribosome sont l'objet d'un contrôle drastique, effectué également par certaines bases des ARNr (FIGURE 17). Celles-ci peuvent en effet former des interactions avec le duplex d'ARN issu de l'appariement entre l'anticodon et le codon (COCHELLA ET AL., 2007).



Ce n'est que si la région de l'ARNt portant l'anticodon est dans une conformation thermodynamiquement stable, et que l'appariement codon/anticodon est correcte, que ces interactions peuvent se former et permettre la suite du processus, par des changements conformationnels au niveau de l'ARNt et du ribosome. (AGRIS ET AL, 2008). Tout défaut de contrôle ou problème d'appariement peut aboutir à un arrêt de la traduction, voire à des décalages du cadre de lecture et à des erreurs qui peuvent être fatales à la cellule. Nous verrons par la suite une des stratégies mises en place par les cellules au niveau des ARNt permettant d'assurer un appariement optimal.

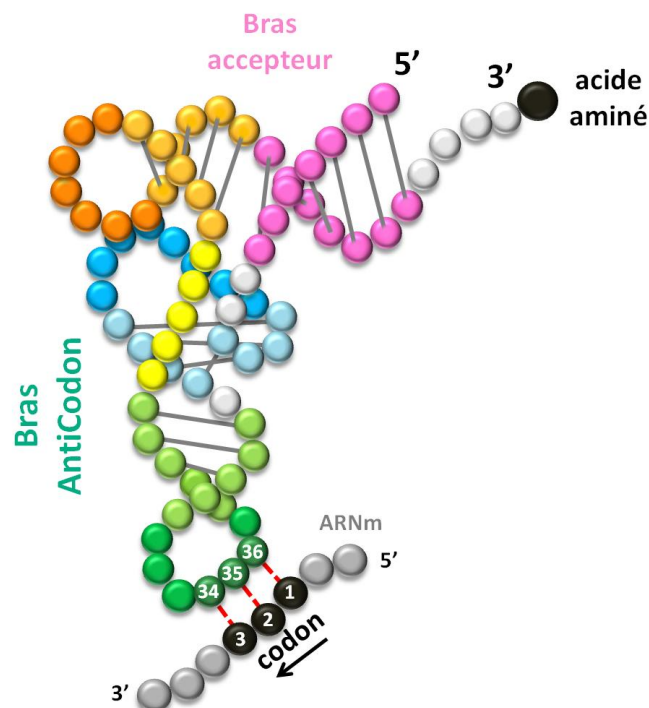
L'orchestration harmonieuse de ces deux grandes étapes fait également intervenir des facteurs d'initiation et d'élongation qui interagissent spécifiquement avec le ribosome, l'ARNm et les ARNt. Enfin, le dernier acte de cette pièce, appelé Terminaison, ne fait pas intervenir d'ARN de transfert : le signal de fin de synthèse du polypeptide, correspondant à l'un des trois codons STOP est reconnu par les facteurs protéiques de terminaison. Le polypeptide est alors libéré, et le ribosome se dissocie de l'ARNm.

Mais avant d'entrer sur scène, à l'image des acteurs d'opéra qui passent entre les mains d'équipes de costumiers et de maquilleurs, les ARNr, les ARNm et les ARNt sont soumis à un processus de maturation spécifique et complexe, assuré par une multitude d'enzymes dont certaines sont universellement conservées, et qui peuvent nous renseigner sur l'histoire évolutive de la machinerie de la traduction. Cette maturation intervient dès la fin de la transcription des différents types d'ARN, et consiste en des clivages et en des additions de divers groupement chimiques. Les ARNt, en tant que matérialisation du code génétique, sont la classe d'ARN qui subissent probablement le plus de transformations avant d'aboutir à une molécule fonctionnelle (PHIZICKY ET AL., 2010, REVUE).



**FIGURE 18 : Représentation de la structure secondaire « en feuille de trèfle » des ARNt**

Chaque cercle représente un nucléotide, dans lequel est indiquée la numérotation standard officielle. Une étoile signifie que le nucléotide est absent chez certains ARNt. Les barres grises représentent les liaisons hydrogènes intramoléculaires. Les numéros des nucléotides de l'anticodon (34, 35 et 36) sont indiqués en blanc. (d'après Sprinzl *et al*, 1998)



**FIGURE 19 : Représentation de la structure tridimensionnelle en « L » classique d'un ARNt**

Les différentes couleurs sont identiques à celles utilisées dans la figure précédente et identifient les différentes régions tiges et boucles (rose= tige ACC, orange = TΨC, bleu=D, vert=AC, jaune= variable). Une portion de l'ARNm est représentée (en gris), appariée avec l'ARNt au niveau du codon (en noir). Les barres rouges représentent les liaisons hydrogènes intermoléculaires établies entre le codon et l'anticodon.



## 2 - L'ARNt

Les ARNt sont de petites molécules d'ARN, d'une taille variant de 78 à 95 nucléotides, dont la structure générale est hautement conservée chez tous les êtres vivants. Chez tous les ARNt, le brin d'ARN est replié en une structure secondaire caractéristique, que l'on appelle en « feuille de trèfle » (*cloverleaf* en anglais) en raison de sa représentation (**FIGURE 18**) (**BERG, BIOCHEMISTRY. 5TH EDITION**).

Cette structure présente quatre régions : une hélice formée des extrémités 5' et 3', appelée tige acceptrice (le « pétiole » du trèfle). L'extrémité 3' est de manière universelle terminée par un triplet de nucléotides ACC, auquel est lié covalement l'un des 22 acides aminés composant les protéines. Les trois autres régions consistent chacune en une structure « tige/boucle » (les « feuilles » du trèfle): une courte hélice double brin, terminée par une boucle simple brin. Elles sont respectivement appelées tige/boucle D, tige/boucle TΨC, et tige/boucle anticodon (AC), car cette dernière contient le triplet de nucléotides s'appariant avec le codon de l'ARNm. Chez certains ARNt, une quatrième région de type boucle, appelée Variable, peut être présente entre les tiges anticodon et TΨC.

Ces différentes tiges/boucles interagissent entre elles, par l'intermédiaire de certains appariements intramoléculaires clés, pour former une structure tridimensionnelle particulière, en forme de L (ou de la lettre grecque gamma majuscule Γ) (**FIGURE 19**). L'une des extrémités de ce L correspond au bras anticodon, l'autre extrémité au bras accepteur, tandis que les tiges/boucles D, TΨC et Variable forment le « coude » de ce L. Cette structure commune détermine la reconnaissance des ARNt par le ribosome, les facteurs d'élongation, ainsi que par différentes enzymes (notamment les enzymes de modification que l'on abordera par la suite), et constitue de ce fait un paramètre clé pour leur fonction.

La nature de l'acide aminé porté au niveau du bras accepteur est directement liée à la séquence de l'anticodon au niveau du bras opposé, cette correspondance détermine les règles établies par le code génétique : à chaque codon, un acide aminé (nous verrons les modalités précises du code génétique par la suite).



L'addition de l'acide aminé au niveau de l'ARNt, appelé amino-acylation, est catalysée par une classe d'enzymes: les amino-acyl-ARNt synthétases (aaRS). Chaque aaRS est spécialisée dans l'ajout d'un seul type d'acide aminé, et reconnaît selon différentes modalités le ou les ARNt qui leur sert de substrat (plusieurs ARNt portant le même acide aminé sont appelés isoaccepteurs). Ces aaRS sont donc des éléments essentiels du système traductionnel : leur activité est la véritable matérialisation du code génétique. Leur étude tant au niveau de leur mécanisme catalytique, de leur modalité de reconnaissance, mais aussi l'étude de leur histoire évolutive constitue un enjeu majeur pour comprendre l'apparition, l'évolution et les modalités du code génétique moderne. (WOESE ET AL, 2000).

Mais l'amino-acylation ne constitue pas l'unique modification chimique post-transcriptionnelle qui affecte les ARNt, et d'autres modifications jouent également un rôle dans les règles subtiles du code génétique...

## **B- Les modifications d'ARNt**

### **1 - Les nucléotides rares, pas si rares que ça !**

$\Psi$ ,  $m^2G$ ,  $mnm^5se^2U$ ,  $cmnm^5s^2Um$ ,  $o^2yW$ ,  $ncm^5U$ ,  $io^6A$ ,  $acp3\Psi$ ,  $ho^5U$ ,  $preQ_0$ ,  $Tm^5s^2U$ ,  $Im$ ,  $C^*$ ,  $imG^{14}$ ,  $s^2Um$ ,  $ms^2hn^6A$ ,  $s^2C$ ,  $D$ ,  $m^1Im$ ,  $Q$ ,  $k^2C$ ,  $inm^5Um$ ,  $OHyW$ ,  $galQ$ ,  $f^5C$ ,  $m^1acp3\Psi$ ,  $t^6A$ ,  $G^+$ ,  $oQ$ ,  $m^2A$ ,  $m4^2Cm$ ,  $chm^5U$ ,  $Gr(p)$ ,  $ms^2t^6A...$

Non, il ne s'agit pas d'un bug de Word, ni d'une crise de nerf sur le clavier de l'ordinateur, encore moins d'une traduction ratée d'un texte chinois par « Google Translate », mais de la liste non exhaustive des nucléosides modifiés connus (référéncés dans la banque de données MODOMICS, MACHNICKA ET AL, VERSION 2013). (FIGURE 20)

En effet, avant même de découvrir le rôle et la structure des ARNt, on identifia au sein des ARN cellulaires des nucléosides atypiques, légèrement différents au niveau chimique des quatre nucléosides canoniques Adénosine, Uridine, Cytidine et Guanosine. On les qualifia donc de « rare », relativement à leur quantité dans la cellule. La Pseudo-uridine, noté  $\Psi$ , fut le premier représentant historique de ces nucléosides atypiques (DAVID & ALLEN, 1957, appelé le « 5<sup>ème</sup> nucléotide »), et est le plus abondant (retrouvé au niveau des ARNr, ARNt et des ARN nucléaires). Sa présence au sein d'un motif caractéristique conservée dans les ARNt est à l'origine du nom de la boucle « T $\Psi$ C ». De même, la présence du nucléoside dihydroxyuridine, noté D, au niveau de l'une des boucles de l'ARNt est à l'origine du nom



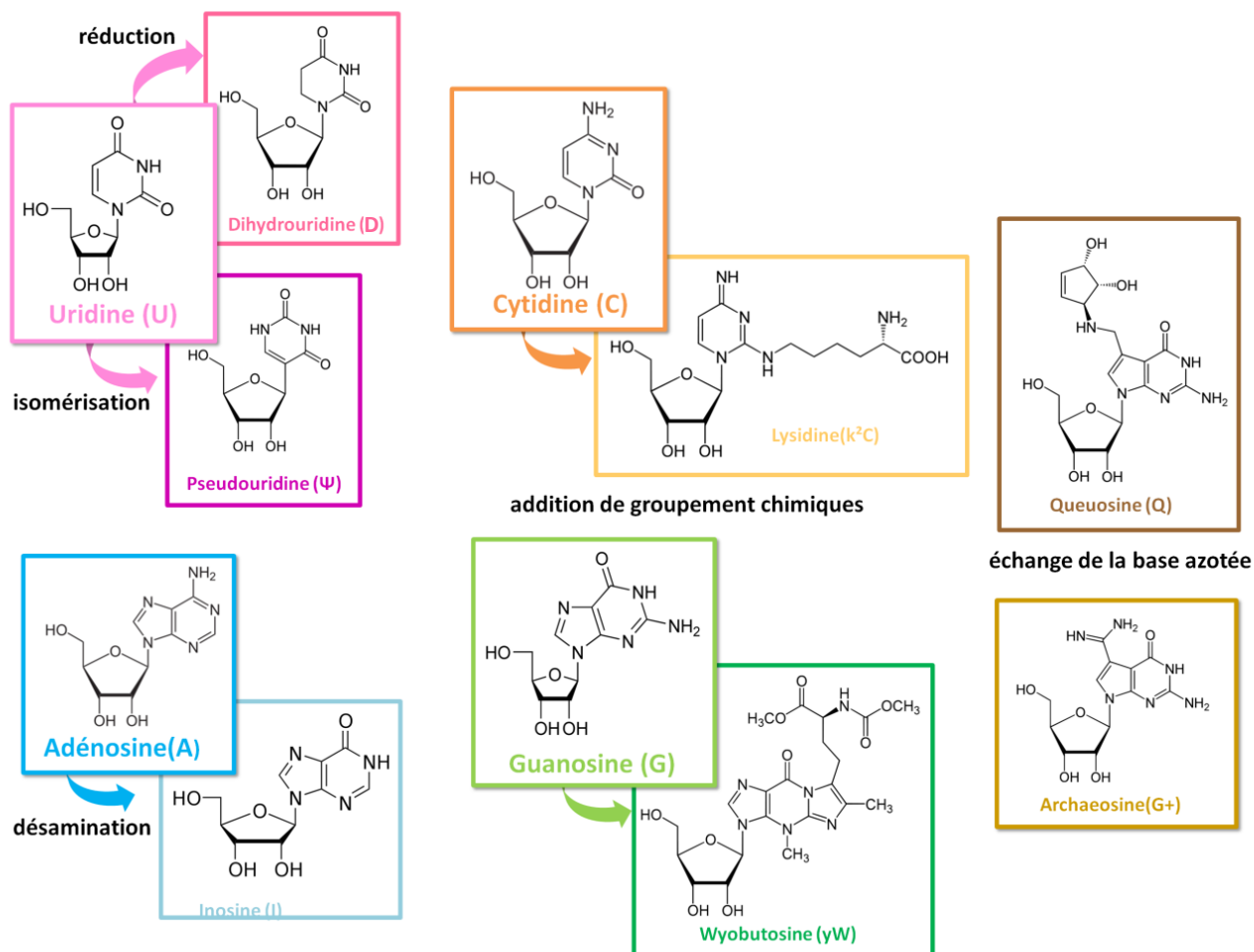
« boucle D ». A l'heure actuelle, environ une centaine de types de nucléosides différents ont été découverts au sein des ARNr, ARNm et ARNt. (FIGURE 20-A) Ces derniers sont les molécules qui présentent la plus grande diversité : 79 de ces nucléosides sont retrouvés au niveau des ARNt, dont 59 de manière spécifique (c'est-à-dire uniquement chez les ARNt) (GROSJEAN&BENNE, MODIFICATION AND EDITING OF RNA, ASM PRESS, 1998). Certains de ces nucléosides ne sont retrouvés que chez certaines espèces, ou groupes d'organismes, alors que d'autres sont commun à deux domaines, voire aux trois grands domaines du vivant (on en recense environ une vingtaine) (FIGURE 20-B). Au bilan, tous les ARNt, chez tous les êtres vivants, présentent au moins un, souvent plusieurs de ces nucléotides dits « rares ». Ainsi, cette fréquence rend obsolète ce qualificatif autrefois employé.

## 2 - Des adeptes du tuning... moléculaire

A l'issue de la transcription, l'ARNt est un oligonucléotide contenant uniquement les nucléosides A, U, G et C. Les nucléosides « atypiques » sont des dérivés post-transcriptionnels de ces quatre nucléosides classiques : on les qualifie donc de nucléosides « modifiés ». Leur formation consiste dans la majorité des cas en une modification chimique plus ou moins complexe du nucléoside de départ (FIGURE 21).

Il peut s'agir d'une isomérisation de la base azotée, comme dans le cas de la pseudo-uridine (notée **Psi**) qui est un C-glycosyl isomère de l'uridine), de la perte d'un groupement chimique comme dans le cas de l'inosine (une adénosine désaminée), ou d'une réduction comme chez la dihydrouridine (la double liaison du cycle aromatique est réduite) (FIGURE 22). Pour de nombreux nucléosides modifiés, il s'agit de l'addition d'un groupement chimique supplémentaire. Ce groupement peut être petit, comme un méthyl (noté m) ajouté sur la base ou le sucre du nucléoside ( $m^1G$ ,  $m^6A$ , Gm, Um...), un groupement thio (soufre) comme pour  $s^2U$ ). Dans ces différents cas, on qualifie ces modifications de « simple ».

Le groupement chimique additionnel peut également consister en une molécule plus complexe, comme pour  $i^6A$  (ajout d'un groupement iso-pentyl),  $t^6A$  et  $g^6A$  (ajout d'un acide aminé carbamoylé), manQ (ajout d'un mannose sur un nucléoside déjà modifié, la queuosine). Dans de nombreux cas, un nucléoside peut porter plusieurs groupements différents, ajoutés successivement, et être chimiquement très éloigné du nucléoside



**FIGURE 22 :** Exemples de modifications chimiques à l'origine de certains nucléosides modifiés

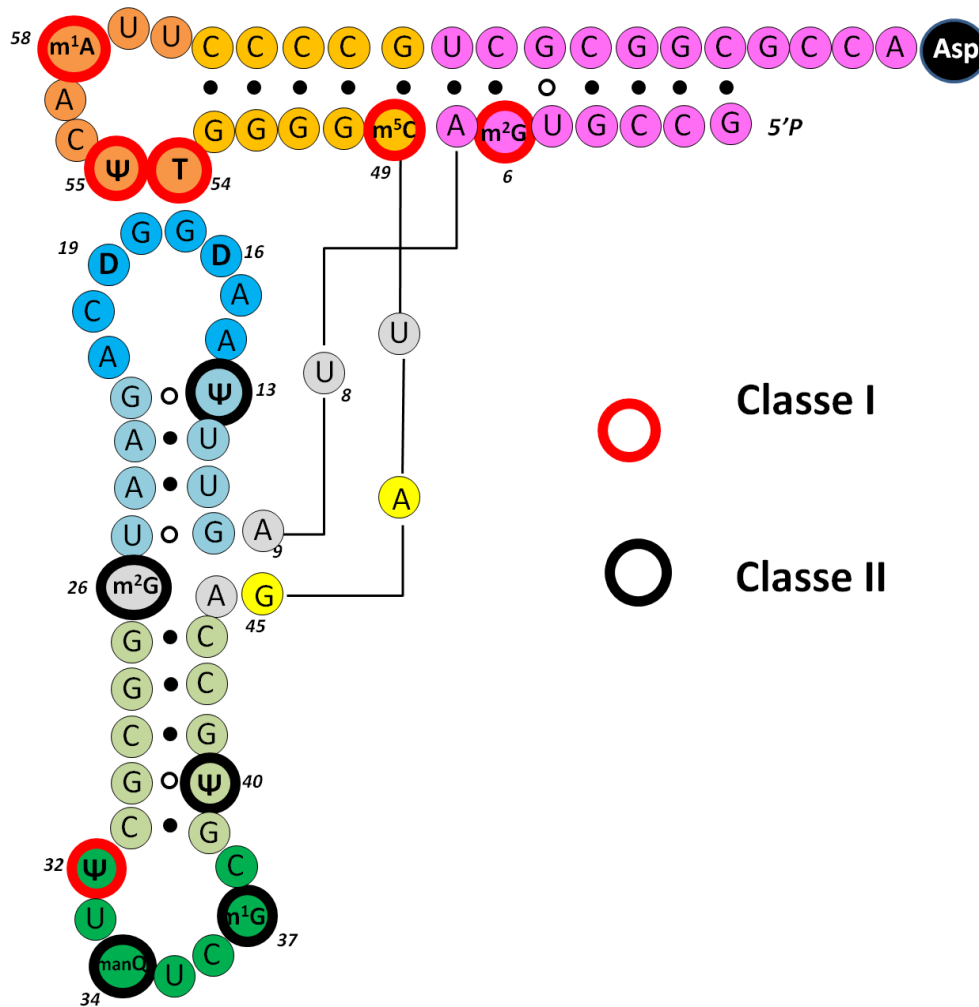
d'origine (comme la wyobutosine). Enfin, un cas particulier concerne la queuosine et l'archaeosine. Chez ces nucléosides, la base azotée d'origine est tout simplement échangée par une molécule hétérocyclique synthétisée par une série d'enzymes (MORRIS ET AL., 2001, PHILLIPS ET AL., 2012).

Ces différents nucléosides, en raison de leur complexité, sont qualifiés « d'hyper-modifiés ». Contrairement aux modifications « simples », que l'on retrouve aussi au niveau des ARNr, et des petits ARN nucléaires (snoARN), les nucléosides hypermodifiés sont caractéristiques des ARNt. Ils représentent 80% des 59 nucléosides modifiés spécifiques aux ARNt. De plus, de manière intéressante, ils sont pour la plupart situés uniquement au niveau de la boucle anticodon, aux positions stratégiques 34 et 37.

### 3 - Les enzymes de modification : un défi pour les biochimistes

Cette incroyable diversité est associée à une panoplie d'enzymes catalysant spécifiquement la formation de ces nucléosides modifiés. A chaque modification correspond souvent une enzyme qui agit directement au niveau de l'ARNt. Dans certains cas, une même modification (par exemple une méthylation sur l'azote d'une guanosine) se situant à différents endroits de l'ARN est catalysée par des enzymes distinctes (isoenzymes). A contrario, il a été découvert des enzymes de modification agissant sur plusieurs positions, voire plusieurs types d'ARN. Par exemple, chez *E. coli*, il a été récemment montré que l'enzyme TrmA catalyse la méthylation de l'uridine à la fois sur l'ARNt (au niveau du bras T) et sur l'ARNtm (RANAEI-SIADAT ET AL., 2013). Les hypermodifications requièrent souvent plusieurs enzymes différentes (par exemple, la formation de wyobutosine implique 5 enzymes différentes, NOMA ET AL., 2006). Il est ainsi difficile d'estimer le nombre d'enzymes de modification dans un organisme donné. Chez *Escherichia coli* par exemple, on estime qu'il existe au moins 45 enzymes de modifications différentes qui sont requises pour synthétiser les 29 types de nucléosides modifiés présents sur l'ensemble des ARNt (BJÖRK, 1995). Le contenu génomique correspondant aux gènes de ces enzymes de modification représenterait environ 4 fois le contenu génomique correspondant aux ARNt eux-mêmes ! (GROSJEAN, MOTORIN, MORIN, CHAP2 « MODIFICATION AND EDITING OF RNA », ASM PRESS, 1998)





**FIGURE 23 :** Exemple de modifications catalysées par des enzymes de classe I et de classe II au sein de l'ARNt<sup>Asp</sup> chez *Xenopus laevis*.

Les cercles rouges correspondent aux nucléosides modifiés catalysés par les enzymes de classe I, insensibles aux perturbations de la structure tertiaire de l'ARNt

Les cercles noirs correspondent aux nucléosides catalysés par les enzymes de classe II, dépendantes de la structure tertiaire de l'ARNt

d'après Constantinesco, modifié de Grosjean *et al.* 1996



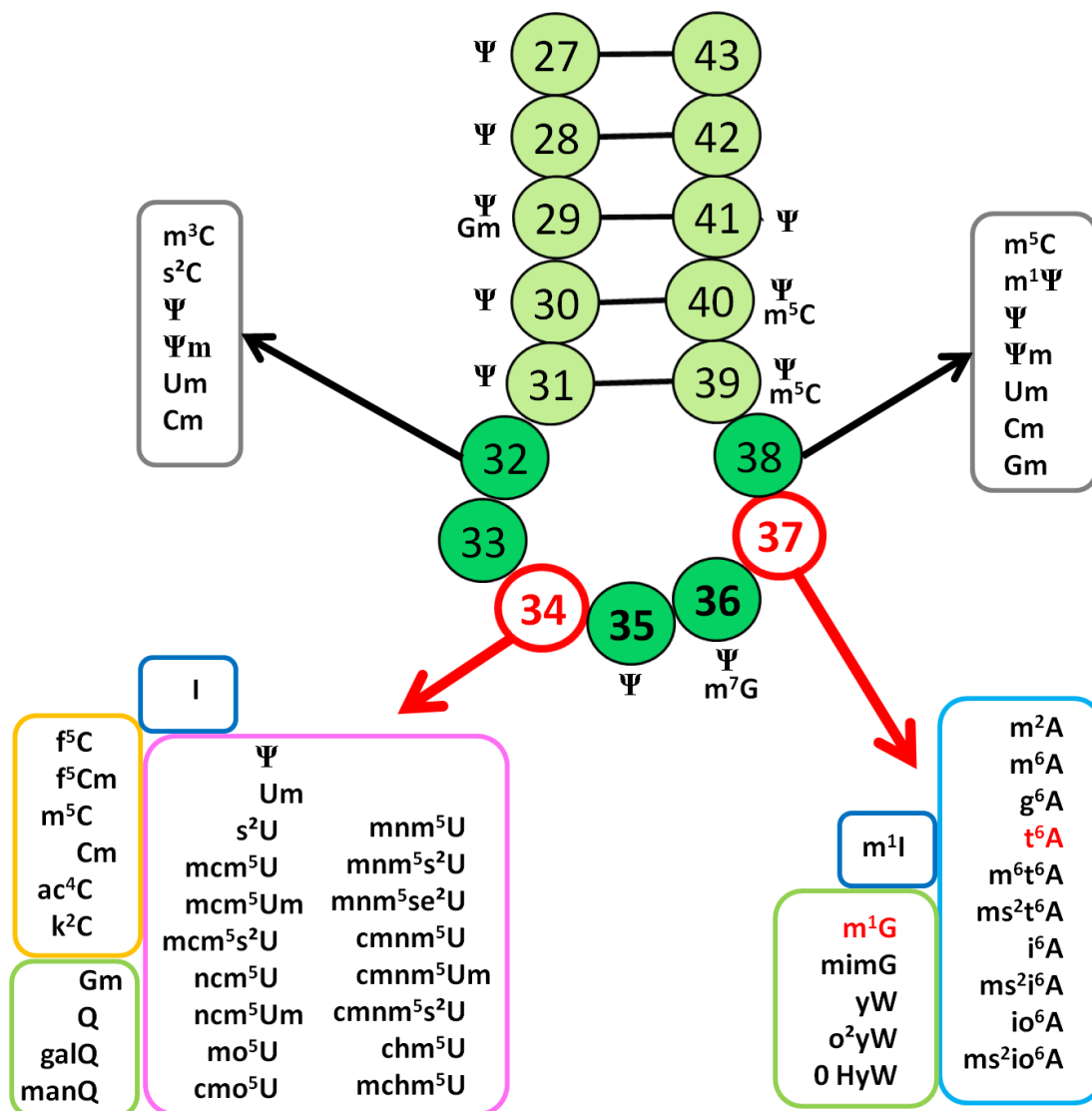
Malgré cette importance manifeste, il existe encore certaines modifications dont les enzymes restent toujours inconnues ou dont la biosynthèse est encore mal comprise : l'une de ces modifications,  $t^6A$ , constituera le sujet de cette thèse.

Il est également difficile d'énoncer des règles générales quant aux mécanismes de modification et au fonctionnement de ces enzymes. L'addition de groupement chimique se traduit par la formation de nouvelles liaisons covalentes. Ainsi, la plupart des enzymes de modification exploitent l'énergie d'hydrolyse de nucléotides (ATP), ou l'utilisation de cofacteurs activés comme le S-Adénosyl-méthionine, utilisé par les méthylases.

Les hypermodifications requièrent notamment de multiples substrats et plusieurs étapes catalytiques successives, assurées par des voies enzymatiques complexes. Les études basées sur la microinjection dans des ovocytes de Xénope de substrats ARNt dont la structure est altérée ont permis de classer les enzymes de modification en deux grands groupes (GROSJEAN ET AL., 1996): les enzymes de classe I, qui ne nécessitent pas d'éléments structuraux pour la reconnaissance du nucléotide à modifier, et les enzymes de classe II, qui requièrent des éléments structuraux au niveau de l'ARNt (FIGURE 23). Les modalités de reconnaissance et de spécificité de chaque enzyme sont donc des paramètres cruciaux pour comprendre les mécanismes de synthèse de chaque nucléoside modifié.

La connaissance des voies enzymatiques de biosynthèse des nucléosides modifiés présente différents intérêts, notamment pour la recherche de nouveaux antibiotiques ciblant spécifiquement les enzymes de modification bactériennes, ou pour la compréhension de maladies génétiques liées à certaines de ces enzymes (nous en verrons un exemple dans la suite du texte concernant la modification  $ms^2t^6A$ ). Enfin, d'un point de vue évolutif, de nombreuses informations peuvent être obtenues par l'étude comparée des systèmes enzymatiques catalysant une même modification chez différents organismes. Nous en verrons un exemple par la suite avec la modification universelle  $m^1G$  en position 37 des ARNt, et l'objet de cette thèse, à savoir la compréhension du système enzymatique de biosynthèse de la modification  $t^6A$ , soulèvera également des questions d'ordre évolutif.

Mais avant de passer à ces questions, il est nécessaire de présenter le rôle fondamental que jouent ces nucléosides modifiés au sein des ARNt.



**FIGURE 24 : Type et localisation des nucléosides modifiés dans la boucle anticodon des ARNt**  
 Hormis les positions 33, 42, 43, toutes les positions sont des sites potentiels de modifications. Presque toutes les hypermodifications sont localisées exclusivement au niveau des positions 34 (appelée la *wobbling base*) et 37 (adjacente à l'anticodon en 3', appelée *dangling base*), représentées en rouge. Les nucléosides  $m^1G$  et  $t^6A$  en positions 37 (indiqués en rouge) sont retrouvés universellement. Extraits de Sprinzl *et al* 1998

## 4 - Rôle des nucléosides modifiés

La présence de groupements, polaires, chargés, hydrophobes, ou stériquement encombrants modifie les propriétés structurales et électrostatiques des nucléosides classiques, leur permettant d'acquérir de nouvelles fonctionnalités.

De nombreuses modifications dans le corps de l'ARNt affectent son repliement et sa stabilité. Par exemple, la présence  $m^1A_9$  chez l'ARNt<sup>Lys</sup> mitochondrial humain à elle seule permet la structure en « feuille de trèfle », et son absence mène à la formation de structures alternatives (HELM ET AL., 1999). Il a été observé une corrélation entre la température de croissance et le nombre de méthylations de l'ARNt, les organismes hyperthermophiles présentant de très nombreuses modifications au niveau du ribose (KOWALAK ET AL, 1994) (McCLOSKEY ET AL, 2002). La méthylation du groupement libre OH du ribose, fonction très réactive à haute température, permettrait d'assurer une protection contre la dégradation. De plus, les modifications permettent d'augmenter la température de fusion de l'ARN. A l'opposé, les organismes psychrophiles présentent de nombreuses Dihydroxiuridines, qui sont supposées augmenter la flexibilité de l'ARNt (DALLGUE ET AL, 1997). Ainsi, la modification d'ARNt est une des stratégies d'adaptation à certaines contraintes environnementales.

D'autre part, certaines modifications à des positions variées affectent spécifiquement l'identité de l'ARNt. Par exemple, l'inosine en position 34 dans l'ARNt<sup>Ile</sup> de *S. cerevisiae* est un déterminant positif pour son aminoacylation (SENGER ET AL, 1997) alors que la lysidine à la position 34 prévient la misacylation de l'ARNt<sup>Ile</sup> par l'enzyme MetRS d'*E. coli* (MURAMATSU ET AL, 1988).

La boucle anticodon est une région de l'ARNt qui concentre de nombreux types de nucléosides modifiées, notamment des hyper-modifications, en particulier au niveau des nucléosides 34 et 37 (FIGURE 24). Les modifications présentes à ces positions influencent de manière très importante les interactions entre codon et anticodon, et constituent ainsi un paramètre modulant la capacité de décodage de l'information génétique par les ARNt, que nous allons maintenant aborder.

POSITION 37 DE  
L'ANTICODON  
(DANGLING BASE)

## DEUXIEME POSITION DU CODON

		U		C		A		G								
<div><div><div>i(o)<sup>6</sup>A</div><div>m<sup>1</sup>G</div><div>W</div><div>t<sup>6</sup>A</div></div><div><div>m<sup>1</sup>G</div><div>m<sup>2</sup>A</div></div><div><div>t<sup>6</sup>A</div><div>m<sup>6</sup>A</div></div><div><div>A</div><div>m<sup>6</sup>A</div><div>m<sup>1</sup>G</div><div>m<sup>2</sup>A</div></div></div> <div>PREMIERE POSITION DU CODON</div>	U	UUU	Phe	Gm <sub>34</sub>	UCU	Ser	xo <sup>5</sup> U <sub>34</sub>	UAU	Tyr	Q	UGU	Cys	U			
		UUC	Phe		UCC	Ser		UAC	Tyr		UGC	Cys		C		
		UUA	Leu Stop		UCA	Ser Stop		UAA	STOP Gln, Leu, Ala		UGA	STOP Cys, Trp, Sec			A	
		UUG	Leu		Cm, f <sup>5</sup> Cm	UCG		Ser	UAG		STOP Gln, Pyl	UGG		Trp		G
	C	CUU	Leu Thr		CCU	Pro	xo <sup>5</sup> U <sub>34</sub>	CAU	His	mnm <sup>5</sup> U <sub>34</sub> , Q	CGU	Arg	I	U		
		CUC	Leu Thr		CCC	Pro		CAC	His		CGC	Arg			C	
		CUA	Leu Thr		CCA	Pro		CAA	Gln		CGA	Arg				A
		CUG	Leu Thr		CCG	Pro		CAG	Gln		CGG	Arg			G	
	A	AUU	Ile	k <sup>2</sup> C, ac <sup>4</sup> C, Cm	ACU	Thr	mo <sup>5</sup> U <sub>34</sub>	AAU	Asn	Q	AGU	Ser		U		
		AUC	Ile		ACC	Thr		AAC	Asn		AGC	Ser			C	
		AUA	Ile Met		ACA	Thr		AAA	Lys Asn		AGA	Arg Ser, Gly, Stop				mcm <sup>5</sup> U <sub>34</sub>
		AUG	Met		ACG	Thr		AAG	Lys		AGG	Arg Ser, Gly, Stop			G	
	G	GUU	Val	xo <sup>5</sup> U <sub>34</sub>	GCU	Ala	xo <sup>5</sup> U <sub>34</sub>	GAU	Asp	s <sup>2</sup> U <sub>34</sub>	GGU	Gly		U		
		GUC	Val		GCC	Ala		GAC	Asp		GGC	Gly			C	
		GUA	Val		GCA	Ala		GAA	Glu		GGA	Gly				A
		GUG	Val		GCG	Ala		GAG	Glu		GGG	Gly			G	
TROISIEME POSITION DU CODON (WOBBLE BASE)																

**FIGURE 25 : Le Code Génétique**

En plus des 64 codons et de leur correspondance, les nucléosides modifiés en position 34 et 37 de l'anticodon des ARNt, qui joue un rôle dans la reconnaissance des codons, sont indiquées à titre indicatif.

Les codons dégénérés 6 fois sont colorés en bleus, ceux dégénérés 4X en rose, dégénérés 3 fois (Ile) en orange, dégénérés 2X en vert, et enfin les codons unique (Met et Trp) colorés en jaune. Les usages alternatifs des codons, utilisés chez certains organismes et les mitochondries, sont indiqués en police réduite de couleur rouge et bleu. Les deux acides aminés Selenocystéine et Pyrrolysine codés chez certains organismes sont indiqués en blanc. Chez la mitochondrie, l'ARNt<sup>Met</sup> reconnaît AUG et AUA (indiqué en jaune). Les modifications de la base 34 (*wobbling*) sont indiquées en vertical dans les différentes boîtes à titre d'exemple. Les modifications de la base 37 sont indiquées sur la gauche du tableau, le nucléoside le plus fréquent (ou ses dérivés) est indiqué en rouge.

X= (c)mnm5, mcm5, x= H, méthyl, acide acétique  
d'après Agris, 2008.

## 5 - Quand la base 34 vacille, le code génétique dégénère...

Le code génétique est dégénéré et semble gelé. Cette constatation a été faite dès les années 1960, lorsque le mode de codage de l'information génétique, sous forme de triplet de nucléotides appelés codons, fut élucidé (NIRENBERG ET AL 1965, SÖLL ET AL 1965). En effet, quatre nucléosides différents (A, U, G, C) sur trois positions (1<sup>ère</sup>, 2<sup>ème</sup>, ou 3<sup>ème</sup> de chaque codon) aboutissent à  $4^3=64$  codons possibles. Or il n'y a que 22 acides aminés naturels dans les protéines. Ainsi, à un acide aminé donné peuvent correspondre plusieurs codons : on parle de dégénérescence. Trois codons sont usuellement utilisés comme signal stop (appelé Ambre\* UAG, Opale UGA et Ocre UAA), et sont reconnus par les facteurs de terminaison. Les 61 codons restants sont appelés sens, et se répartissent en plusieurs groupes de codons « synonymes », c'est-à-dire codant un même acide aminé (FIGURE 25). Lorsqu'elles furent décryptées, ces règles semblaient universelles et applicables à tous les êtres vivants : Francis Crick estima donc qu'elles devaient avoir été établies aléatoirement (par accident), très tôt dans l'évolution, et n'avait depuis subi aucun changement, comme gelé : c'est l'hypothèse du « *frozen accident* ». (CRICK, 1968).

Pour expliquer la dégénérescence au niveau mécanistique, F.Crick émit également l'hypothèse que les deux premières bases du duplex codon/anticodon devaient s'apparier strictement selon une interaction Watson-Crick, (A=U, G=C), alors que la troisième base du codon devait s'apparier de façon plus lâche avec la base 34 de l'anticodon (G pouvant alors s'apparier avec U): c'est l'hypothèse du *wobbling* (qui se traduit par « vaciller » en français) (CRICK, 1966). Cette hypothèse a été confirmée par la suite : en effet on constate que la position 34 de l'anticodon (appelé *wobbling base*) admet des appariements moins stricts, permettant à un même ARNt de lire plusieurs codons. Ainsi, chez *E. coli*, on dénombre seulement 47 espèces différentes d'ARNt qui suffisent à décoder l'ensemble du code génétique. Certaines espèces ont même un nombre encore plus réduits d'ARNt pour assurer ce décodage (par exemple, certains Mycoplasmes n'ont que 23 ARNt différents).

---

\* le nom de Ambre provient d'un étudiant en thèse, Harris Bernstein (ambre, en allemand), qui, en échange d'un long et fastidieux travail d'isolation de plus de 2000 mutants du phage T4, fut autorisé à nommer les mutants découverts. Les mutants « ambre » du phage T4 possèdent des mutations dans les codons stop UAG).

Hypothèse du Wobble		Hypothèse du Wobble « modifiée »	
Anticodon N34	Codon N3	Anticodon N34	Codon N3
G	C,U	G	C,U
C	G	C	G
I	U,C,A	I	U,C,A
U	A,G	xm <sup>5</sup> U*	G
		s <sup>2</sup> U**	A,G
		xo <sup>5</sup> U***	A,G,U, (C)

\*= tous les dérivés du 5-methylene uridine, à l'exception du 2-thio

\*\*= tous les dérivés du 2-thiouridine

\*\*\*= tous les dérivés du 5-oxyuridine

**FIGURE 26 : L'hypothèse du Wobbling et sa version « modifiée ».**

Tableau indiquant les appariements prévus par l'hypothèse originale prévue par F. Crick et sa version actualisée prenant en compte les modifications de la base 34. D'après Agris, 2007.

A l'origine, le modèle de *wobbling* prenait déjà en compte, en plus des appariements non canoniques G:U et U:G, la présence de l'inosine (un dérivé de l'adénosine en position 34 découvert en 1966) pouvant s'apparier à U, C ou A (FIGURE 26). Cependant, ce modèle ne permet pas d'expliquer en totalité les différentes modalités de reconnaissance des codons par les ARNt observées expérimentalement, et on découvrit finalement le rôle essentiel des modifications d'ARNt dans le phénomène de *wobbling*.

L'une des positions les plus modifiées dans les ARNt est en effet la *wobbling base* (position 34) (cf FIGURE 24). On compte plus d'une trentaine de modification qui peuvent l'affecter, dont de nombreuses hypermodifications, notamment lorsqu'il s'agit d'une uridine. Ces modifications, en changeant les propriétés des nucléotides, permettent de contrôler très précisément l'appariement avec le codon, en restreignant le phénomène de *wobbling* ou au contraire en augmentant ses possibilités. On parle alors de « *expanded wobbling* » (« vacillement étendu ») (FIGURE 26). Certaines modifications ne changent pas les propriétés de reconnaissance, mais augmentent la stabilité de l'appariement et contribuent à la fidélité de lecture. Ainsi, la présence des modifications est maintenant pris en compte dans la lecture du code génétique « standard », et la caractérisation de l'impact de chacun des diverses modifications sur cette base 34 est essentielle pour comprendre les modalités complexes qui permettent aux cellules de décoder correctement l'information génétique (AGRIS 2007, REVUE).

D'un point de vue évolutif, cette position retient également une attention particulière, car les modifications à son niveau ont été des outils permettant aux cellules d'exploiter les potentialités et possibilités du code génétique. L'identification de certains organismes (notamment les Mycoplasmes, les mitochondries) ayant réassigné certains codon, et la compréhension du rôle des modifications de la base 34 ont montré que le code génétique, que l'on considérait comme gelé depuis des temps immémoriaux, ne l'est en fait probablement pas (GROSJEAN, CRECY-LAGARD, MARCK, 2010)





## 6 - Dangling Queen : la base 37

Au sein de la boucle anticodon, on trouve un autre nucléoside stratégique: il s'agit de la position 37. Adjacent à l'anticodon en 3', le nucléoside positionné à ce niveau est (presque) toujours une purine : A ou G. Des études structurales montrent que cette purine 37 change de géométrie lors de la liaison de l'ARNt au sein du site A du ribosome (contrairement au reste du domaine anticodon), et se positionne comme une plate-forme située au dessus du « complexe » formé par la troisième base de l'anticodon et la première base du codon (AGRIS, 2008). Elle est ainsi qualifiée de « *dangling base* », (en français, se balancer) et joue un rôle structural de première importance. Tout comme la position 34, le nucléoside 37 de l'ARNt présente une grande variété de modifications (cf FIGURE 24). D'une manière générale, la présence d'un groupement additionnel sur la base 37 participe à la structuration de la boucle anticodon, augmente la force d'empilement (*stacking*) des bases de l'anticodon et stabilise l'appariement codon/anticodon. De ce fait, elle permet à l'ARNt de se lier correctement à l'ARNm au sein du ribosome, et empêche des décalages du cadre de lecture : elle joue donc un rôle capital dans l'efficacité et la fidélité de la traduction.

La nature du nucléoside modifié en position 37 est directement liée à la composition de l'anticodon adjacent, notamment de la troisième base 36 (cf FIGURE 25). Une adénosine en position 36 est généralement associée à la présence d'une isopentenyl-adénosine ( $i^6A$ ) en position 37. Une guanosine en position 36 est souvent associée à une méthylguanosine  $m^1G$  en position 37. En raison de cette position stratégique, chacune de ces deux modifications d'ARNt a fait l'objet d'études intensives afin de comprendre leur rôle au niveau de l'ARNt et les mécanismes de leur biosynthèse. Je présenterai donc brièvement certaines données intéressantes issues de l'étude de  $i^6A_{37}$  et  $m^1G_{37}$ , qui pourront éclairer certains aspects des travaux présentés par la suite

.

### **a) Le nucléoside modifié i<sup>6</sup>A<sup>37</sup>**

L'isopentyl-adénosine, noté i<sup>6</sup>A, est un nucléoside modifié présent uniquement en position 37, chez les Bactéries, les Eucaryotes mais aussi les mitochondries et les chloroplastes. L'isopentenyl est un groupement stériquement encombrant et hydrophobe. Présente au niveau des ARNt décodant les codons UNN (N étant n'importe lequel des 4 nucléosides), cette modification est catalysée par une famille de protéines orthologues, appelée MiaA chez les bactéries, Mod5 chez la levure (DIHANICH ET AL., 1987) et TRIT1 chez l'Homme. Les Archées ne semblent pas posséder cette enzyme, et leurs ARNt présentent m<sup>1</sup>G à la place de i<sup>6</sup>A.

Les mutants *miaA* chez *E. coli* et *S. enterica* présentent de multiples défauts traductionnels : une diminution de l'efficacité et de la fidélité de lecture. Ces mutants sont viables mais présentent des phénotypes pléiotropiques: une croissance réduite, une sensibilité altérée aux analogues d'acides aminés, d'importantes perturbations métaboliques. Ainsi, la présence de i<sup>6</sup>A assure une traduction fidèle et efficace, et permet l'expression correcte de différentes protéines impliquées dans de nombreuses voies métaboliques.

De manière intéressante, la protéine Mod5 chez *S. cerevisiae* est exportée à la fois dans le cytoplasme et dans la mitochondrie, par l'usage d'un codon d'initiation alternatif lors de son expression (BOGUTA ET AL., 1994). Cette enzyme modifie donc à la fois les ARNt cytoplasmiques et mitochondriaux. De plus, alors que chez les Bactéries tous les ARNt décodant les codons UNN portent i<sup>6</sup>A, chez les Eucaryotes, seul un sous-ensemble de ces ARNt est modifié. Ainsi, les modalités de sélectivité et de reconnaissance des ARNt par cette famille d'enzymes semblent complexes et variables selon les organismes d'origine, et ont donc été très étudiées. Récemment, la comparaison des structures de MiaA (de *E. coli*) et Mod5 (de *S. cerevisiae*) avec l'ARNt a permis d'élucider ces différences au niveau mécanistique et structurale (LAMICHHANE ET AL., 2011, SEIF ET AL., 2009, ZHOU ET AL., 2008). Ainsi, ces deux enzymes pourtant orthologues ne sont pas dépendantes des mêmes déterminants moléculaires et ne fonctionnent pas selon les mêmes modalités.

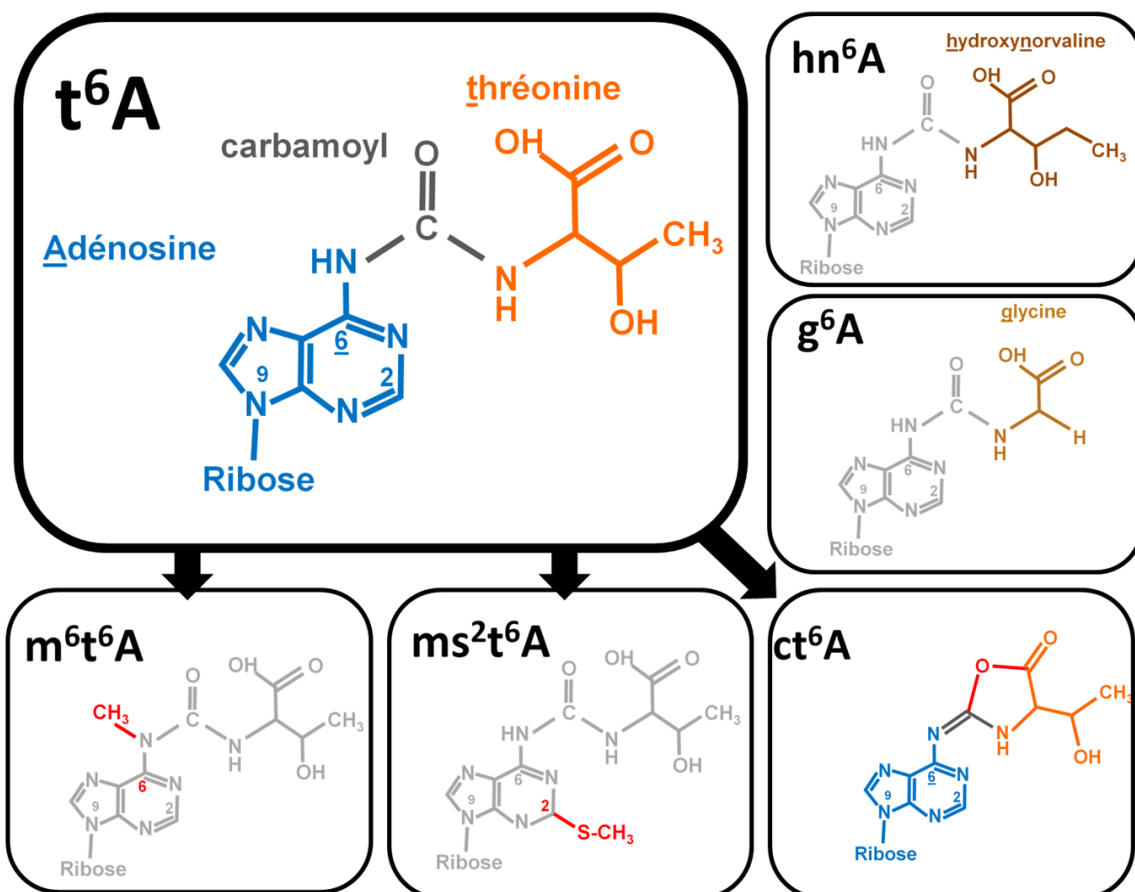


### **b) La modification universelle m<sup>1</sup>G<sup>37</sup> :**

Le nucléoside modifié m<sup>1</sup>G<sub>37</sub> est retrouvé universellement chez tous les organismes (Bactéries, Archées et Eucaryotes), ainsi que chez les mitochondries et les chloroplastes. Présente sur presque tous les ARNt portant un G en position 36 (décodant les codons CNN), mais aussi de manière fréquente sur les ARNt décodant les codons UNN (notamment chez les Archées) elle joue, tout comme i<sup>6</sup>A, un rôle important en augmentant la fidélité de décodage du ribosome (LEE ET AL., 2007). Sa synthèse est catalysée par une méthyltransférase. Chez les Bactéries, cette enzyme est appelée TrmD alors que chez les Archées et les Eucaryotes, elle est appelée Trm5. (GOTO-ITO 2008, 2009). Tout comme Mod5, la protéine Trm5 est présente chez la Levure à la fois dans le cytoplasme et la mitochondrie, la version mitochondriale possédant un peptide d'adressage spécifique probablement issu d'une initiation alternative (LEE ET AL., 2007).

De manière intéressante, les deux familles de protéines TrmD et Trm5 ne sont pas orthologues : elles présentent des repliements distincts, des modalités de reconnaissance de l'ARNt et également un mécanisme catalytique différent (CHRISTANT ET HOU 2007). Ainsi, la présence de m<sup>1</sup>G au niveau de la position 37 pourrait illustrer un cas de convergence évolutive : cette modification pourrait être apparue après la divergence des lignées Bactéries et Archées/Eucaryotes, et donc ne pas avoir été présente chez LUCA. Ainsi, on constate que l'étude des voies enzymatiques des modifications d'ARNt universelles peut nous fournir des renseignements intéressants sur l'histoire évolutive des lignées cellulaires. De plus, les différences entre ces deux familles analogues permettent la recherche de nouveaux antibiotiques, par la caractérisation de molécules ciblant spécifiquement l'enzyme bactérienne sans affecter les versions archéenne/eucaryote (LAHOUD ET AL., 2011).

**Au bilan, les différentes études concernant ces deux modifications de la base 37 ont montré leur rôle essentiel dans la qualité de la traduction. La compréhension de leur voie de biosynthèse respective chez différents organismes a fourni de nombreuses surprises, tant au niveau mécanistique qu'évolutif. Une autre modification au niveau de la position 37 présente des caractéristiques similaires à ces deux nucléosides modifiés : tout comme i<sup>6</sup>A, il s'agit de l'addition d'un groupement chimique stériquement encombrant. Tout comme m<sup>1</sup>G, elle est universelle, présente chez tous les organismes. Tout comme ces deux modifications, on la retrouve sur une classe spécifique d'ARNt : ceux arborant un U en position 36. Cependant, à la différence de m<sup>1</sup>G et i<sup>6</sup>A, les enzymes catalysant sa formation sont restées mystérieuses depuis 40 ans, jusqu'à très récemment... Revenons donc sur la découverte, la caractérisation et l'étude de ce nucléoside modifié : la thréonylcarbamoyladénosine, ou t<sup>6</sup>A.**



**FIGURE 27:** Structure chimique du nucléoside N<sup>6</sup>-thréonycarbamoyladénosine (t<sup>6</sup>A) et de ses variants

L'adénosine est indiquée en bleu, la numérotation officielle de la base azotée est indiquée sur certains atomes.

Le groupement carbamoyl est indiqué en gris, et le groupement thréonyl est indiqué en orange.

Les dérivés méthyl (m) et thiométhyl (ms) ou cyclique (c) de t<sup>6</sup>A, retrouvés chez certains organismes, sont représentés dans les encarts inférieurs, les groupements additionnels sont indiqués en rouge

Les nucléosides similaires à  $t^6A$  :  $g^6A$  et  $hn^6A$ , où l'acide aminé greffé au groupement carbamoyl diffère, sont représentés dans les encarts de droite. d'après Sprinzl *et al.*, 1998, Miyauchi *et al.*, 2012

## C – Le nucléoside modifié t<sup>6</sup>A

### 1 - Une modification universelle - découverte de t<sup>6</sup>A

Le nucléoside N-(adenu-6-yl-carbamoyl)threonine, ou t<sup>6</sup>A, fut découvert en 1969 au sein d'ARNt issus de différents organismes (CHHEDDA ET AL., 1969, SCHWEIZER ET AL., 1969). Chimiquement, il s'agit d'une adénosine sur laquelle est greffé, sur l'atome d'azote 6 de la purine, un groupement urée (carbamoyl) suivi d'une thréonine (FIGURE 27).

Ces premières études révélèrent que t<sup>6</sup>A est présente à la fois chez les Bactéries (*Escherichia coli*) mais aussi chez différents Eucaryotes (champignon : *Torula*, *Rizopongo*, plantes : *Triticum*, mammifères : *Rattus*, *Homo*) (CHHEDDA ET AL 1969) (CUNNINGHAM ET AL., 1974). Après la découverte des Archées, la présence de t<sup>6</sup>A fut étendue à ce nouveau domaine, lors du séquençage du premier ARNt d'Archée (tRNA<sup>met</sup> de *Thermoplasma acidophilum*) (KILPATRICK ET AL., 1981). L'identification des modifications d'ARNt chez une grande variété d'organismes, facilitée par l'évolution des techniques (LC-MS/MS), et la mise en place de banques de données régulièrement actualisées (SPRINTZ ET AL, MODOMICS) complétèrent ces travaux au fil des ans, permettant de fournir un aperçu global de la répartition de ce nucléoside modifié au sein du monde vivant.

t<sup>6</sup>A apparaît alors comme un nucléoside universel, présent systématiquement chez tous les organismes dont les ARNt ont été séquencés, y compris chez les organismes aux génomes très réduits que sont les Mycoplasmes, ainsi qu'au niveau des ARNt des plastes et des mitochondries (GROSJEAN ET AL., 1995). Cette universalité fait de t<sup>6</sup>A un candidat probable pour faire partie des modifications d'ARNt déjà présentes chez LUCA.

La localisation exacte de ce nucléoside au sein des ARNt fut identifiée dès sa découverte : t<sup>6</sup>A se trouve exclusivement en position 3' de l'anticodon, correspondant à la position 37 (TAKEMURA 1969) et présent systématiquement et exclusivement au niveau des ARNt décodant les codons de la forme ANN (ISHIKURA ET AL, 1969, POWERS ET AL, 1972, SPRINTZ ET AL 2005) (FIGURE 28). Une exception concerne toutefois l'ARNt<sup>Met</sup> initiateur des Bactéries, des Archées, des plastes et des mitochondries, qui décodent le codon AUG mais ne présentent pas t<sup>6</sup>A en position 37.

		DEUXIEME POSITION											
		U		C		A		G					
PERMIERE POSITION	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U C A G	TROISIEME POSITION (WOBBLE)		
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys				
		UUA	Leu	UCA	Ser	UAA	STOP	UGA	STOP				
		UUG	Leu	UCG	Ser	UAG	STOP	UGG	Trp				
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U C A G			
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg				
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg				
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg				
	A	AUU	Ile	ACU	Thr	m <sup>6</sup> t <sup>6</sup> A (B)	AAU	Asn	AGU	Ser		U C A G	
		AUC	Ile	ACC	Thr		AAC	Asn	AGC	Ser			
		AUA	Ile	ACA	Thr	AAA	Lys	m <sup>5</sup> 2 <sup>6</sup> t <sup>6</sup> A (E, B, A)	AGA	Arg			
		AUG	Met	ACG	Thr	AAG	Lys		AGG	Arg			
			*Met <sup>i</sup> (B, A, O)										
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U C A G			
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly				
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly				
GUG		Val	GCG	Ala	GAG	Glu	GGG	Gly					

**FIGURE 28 :** t<sup>6</sup>A et ses variants sont présents sur les ARNt reconnaissant les codons de la forme ANN

Le code génétique est représenté sous sa forme classique : codon/acide aminé correspondant

Les cases colorées en rouge correspondent aux codons qui sont reconnus par des ARNt portant t<sup>6</sup>A<sub>37</sub>. La case gris foncé correspond à l'ARNt<sup>Met</sup> initiateur des Bactéries (B), Archées (A) et Organites (O), qui ne porte pas t<sup>6</sup>A. La case rose indique la présence, chez les Bactéries, de m<sup>6</sup>t<sup>6</sup>A au niveau des ARNt<sup>Thr</sup>. La case orange indique la présence, chez les trois domaines du vivant, de ms<sup>2</sup>t<sup>6</sup>A au niveau des ARNt<sup>Lys</sup>

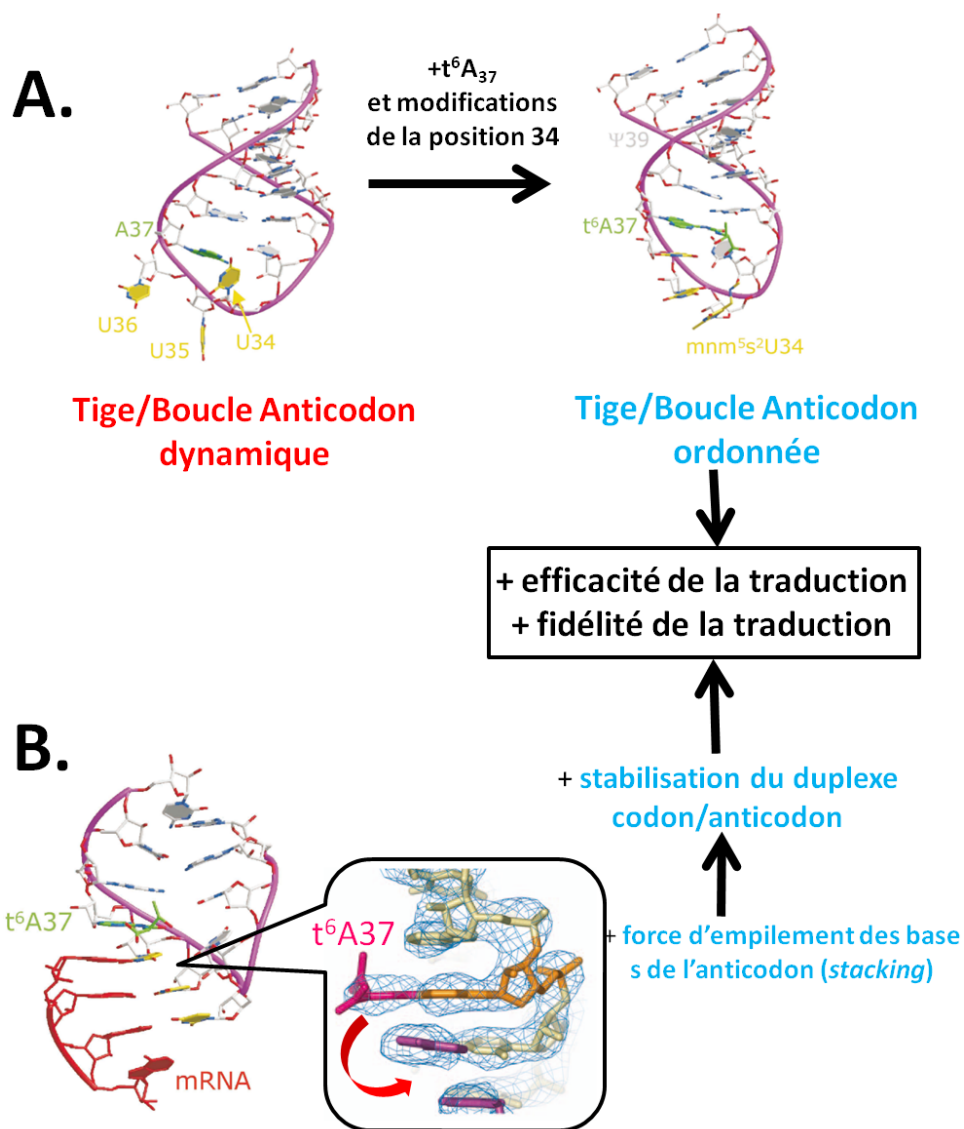


Des variants de ce nucléoside ont également été identifiés et caractérisés (**FIGURE 27 ET 28**): g<sup>6</sup>A (une glycine au lieu de thréonine, identifié chez la Levure (SCHEIWZER ET AL., 1970), hn<sup>6</sup>A (une hydroxynorvaline, trouvé chez certaines Bactéries et Archées thermophiles (REDDY ET AL., 1992), ainsi que leurs dérivés : m<sup>6</sup>t<sup>6</sup>A (chez *E. coli*, sur deux ARNt<sup>Thr</sup>, KIMURA, 1972) et ms<sup>2</sup>t<sup>6</sup>A (dérivé thiométhyl de t<sup>6</sup>A, présent sur l'ARNt<sup>Lys</sup> dans les trois domaines du vivant SPRINZL ET AL., 2005), ms<sup>2</sup>hn<sup>6</sup>A. Récemment, il a été décrit chez plusieurs organismes une forme cyclique de t<sup>6</sup>A : ct<sup>6</sup>A (MIYAUCHI ET AL., 2013). Les auteurs de cette étude suggèrent que la forme non cyclique de t<sup>6</sup>A, observée jusqu'alors dans toutes les études, pourrait être un artefact produit par les conditions d'isolation des ARNt utilisées couramment. Cette forme semble présenter une distribution particulière dans le monde vivant : elle a été détectée chez *E.coli*, *B.subtilis*, *S.cerevisiae* et *S.pombe*, *Spinacia oleracea* (l'épinard) alors que chez l'Homme, la souris, les Archées *H.volcanii*, *T.acidophilum* et les Bactéries *M.mobile* et *A. ferrooxidans* ne semblent pas la posséder, et sa formation implique une protéine de la famille MoeB/ThiF, des enzymes *E1 ubiquitine-activating like* retrouvées dans les trois domaines du vivant.

**Qu'il s'agisse d'une forme dérivée ou de la forme « classique » t<sup>6</sup>A, la compréhension du rôle biologique de ce nucléoside modifié devint très tôt un enjeu majeur, en raison de son universalité associée à sa position clé au niveau des ARNt.**

## 2 - Caractérisation du rôle biologique de t<sup>6</sup>A

La structure de t<sup>6</sup>A, résolue en 1974, montre que la présence du groupement carbamoyl-thréonine empêche l'adénine de pouvoir former un appariement de type Watson-Crick. Cela permet de suggérer que cette modification pourrait intervenir d'une part, dans la conformation de la boucle anticodon et d'autre part jouer un rôle dans la lecture correcte des codons (PARTHASARATHY ET AL., 1974). De nombreuses études biochimiques (MILLER ET AL., 1976, WEISSENBAACH ET AL. 1981, REDDY ET AL. 1981) basées sur des expériences d'interactions entre l'ARNt et l'anticodon selon différents contextes (en absence ou en présence de ribosome), et des études structurales, basées sur la résolution de structures de boucle anticodon dépourvue ou non de t<sup>6</sup>A, vinrent compléter ces travaux initiaux (**FIGURE 29**). (STUART ET AL., 2000, YARAN ET AL., 2000, SUNDARAM ET AL., 2000, MURPHY ET AL., 2005, LECRINIER ET AL., 2006)



**FIGURE 29 : Illustration du rôle structural de t<sup>6</sup>A au niveau de la boucle anticodon de l'ARNt**

A. Comparaison structurale de la tige/boucle anticodon de l'ARNt<sup>Lys(UUU)</sup> non modifiée (à gauche) ou modifiée (à droite). t<sup>6</sup>A<sub>37</sub> (en vert) et mnm<sup>5</sup>s<sup>2</sup>U<sub>34</sub> (en jaune) et Ψ<sub>39</sub> (en gris) permettent de structurer la boucle et de diminuer son entropie.

B. Structure de la boucle anticodon appariée à un ARNm au niveau du site A du ribosome. La base 37, portant t<sup>6</sup>A, se positionne comme une plate-forme au dessus de duplex codon anticodon et augmente les forces d'empilement des bases (*stacking*). D'après Lecrinier *et al.*, 2006.

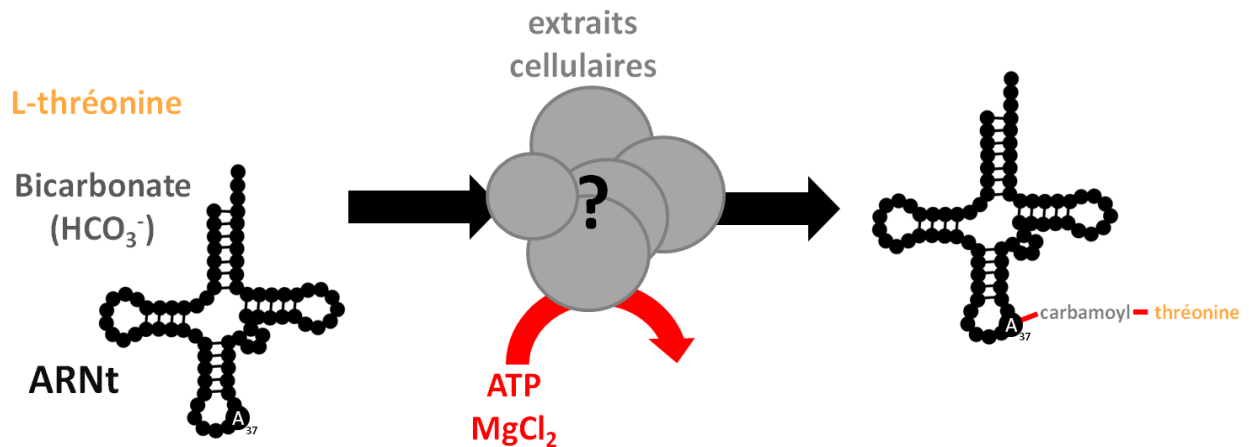
La présence de t<sup>6</sup>A structure la boucle anticodon en empêchant l'appariement entre U33-A37, et réduit son entropie (REVUE, AGRIS ET AL., 2006). t<sup>6</sup>A ordonne également le positionnement des bases de l'anticodon, et augmente la force d'empilement au niveau de ces bases lors de l'interaction avec l'ARNm. Ainsi, la présence de t<sup>6</sup>A stabilise fortement le duplex codon/anticodon, ce qui constitue un paramètre critique pour l'efficacité de la traduction. La présence de ce groupement polaire carbamoylthréonine empêcherait également le décalage de cadre de lecture, et participerait donc à la fidélité de la traduction

Enfin, la présence de t<sup>6</sup>A influencerait également les capacités de décodage de certains ARNt. Ainsi, chez l'ARNt<sup>Lys(UUU)</sup>, qui a été très étudié au niveau structural, la présence simultanée de t<sup>6</sup>A en position 37 et de mcm<sup>5</sup>U en position 34 est nécessaire pour que l'ARNt puisse décoder à la fois les codons AAG et AAA : t<sup>6</sup>A favorise ainsi le phénomène de *wobbling* (YARAN ET AL., 2000)

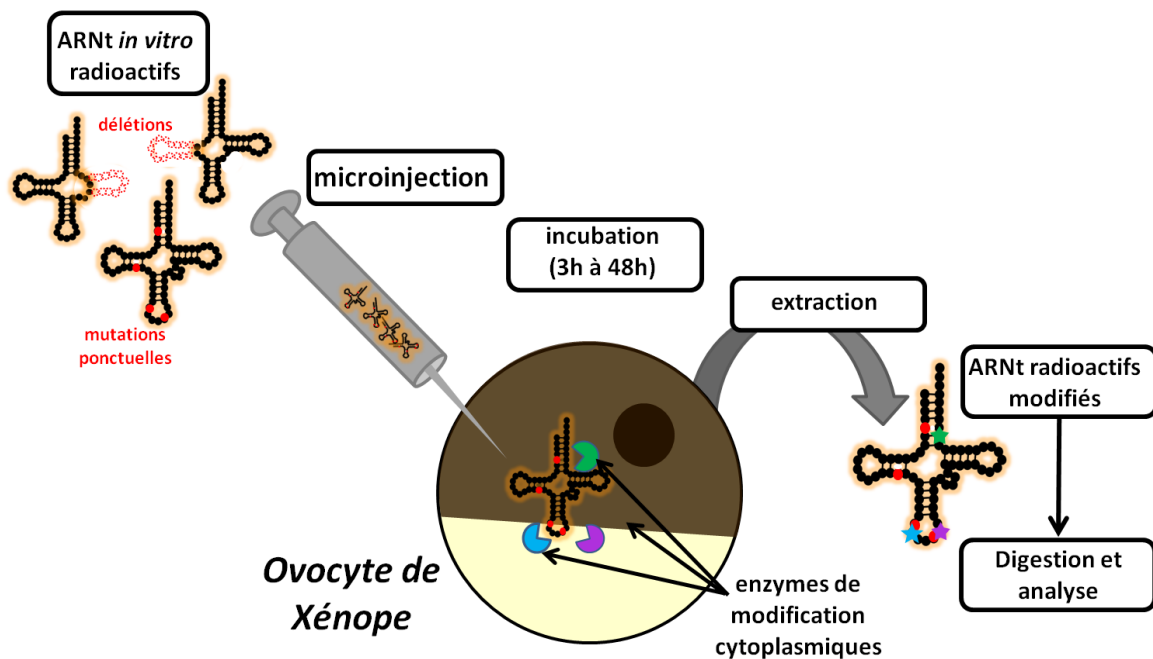
Chez l'ARNt<sup>Met</sup> initiateur Eucaryote en revanche, la présence de t<sup>6</sup>A aurait pour rôle de restreindre l'initiation au seul codon AUG, contrairement aux Bactéries et aux Archées dont l'ARNt<sup>Met</sup> initiateur, dépourvu de t<sup>6</sup>A, peut se s'apparier à GUG et UUG (DUBE ET AL. 1968, LECRINIER ET AL., 2006).

**Ainsi, la présence de t<sup>6</sup>A au sein de certains ARNt semble être l'une des clés qui permet aux organismes d'effectuer une traduction fidèle et efficace. Ce rôle essentiel permettrait d'expliquer la présence universelle de ce nucléoside modifié chez tous les organismes vivants.**

**Mais comme toute étude *in vitro*, une validation du modèle à l'aide de données *in vivo*, par l'étude de mutants dépourvus de ce nucléoside, est indispensable. Pour confirmer le réel impact de t<sup>6</sup>A sur la traduction, il faut pouvoir identifier la ou les enzyme(s) catalysant sa formation, et donc comprendre sa voie de biosynthèse. De plus, la comparaison des machineries enzymatiques chez les Bactéries, les Eucaryotes, les Archées et les organites permettrait d'un point de vue évolutif de déterminer si t<sup>6</sup>A pouvait être présente chez LUCA, et donc donner des renseignements sur la qualité du système traductionnel ancestral et son évolution.**



**FIGURE 30 :** Illustration des connaissances concernant la biosynthèse de t<sup>6</sup>A en 1975, montrant les différents précurseurs identifiés à l'époque.



**FIGURE 31 :** Principe de la technique de micro-injection d'ARNt dans des ovocytes de Xénope permettant l'étude des modifications d'ARNt *in cellulo*.

Des ARNt transcrits *in vitro* et radiomarqués, pouvant porter des mutations ponctuels ou des délétions, sont micro-injectés dans le cytoplasme d'ovocytes de Xénope non fécondés. Après incubation, les ARNt sont extraits puis digérés afin d'analyser la présence de nucléoside modifiés radioactifs, traduisant l'action d'enzymes de modifications cytoplasmiques sur les substrats injectés. D'après Grosjean *et al*, 1987.

### 3 - Caractérisation de la biosynthèse de t<sup>6</sup>A

#### a) Les précurseurs de t<sup>6</sup>A :

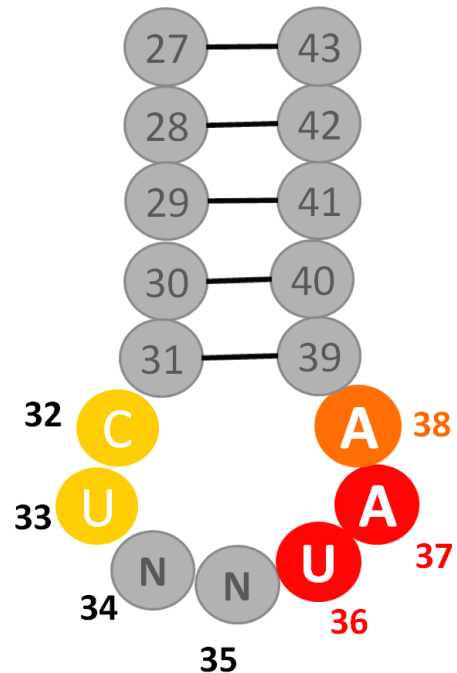
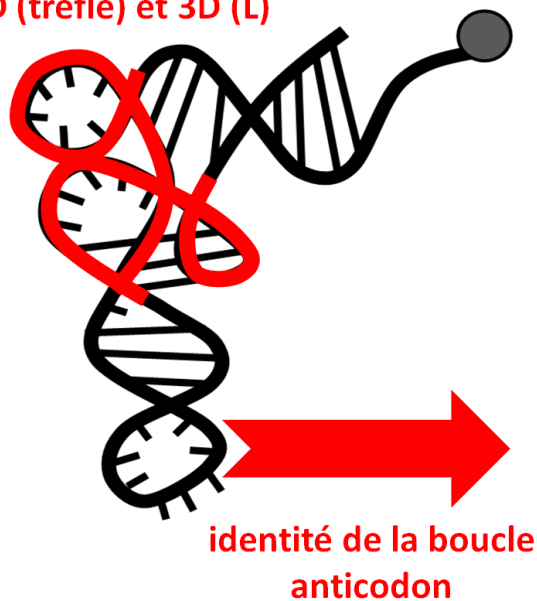
Dès 1972, plusieurs travaux identifièrent les substrats précurseurs du groupement chimique thréonyl-carbamoyl porté par l'adénosine 37 (POWERS 1972, CHHEDA 1972, ELKINS AND KELLER 1974, KORNER AND SOLL 1974). Ces premières études, réalisées chez le rat et chez *E. coli*, montrèrent que t<sup>6</sup>A est synthétisée à partir de thréonine libre, de carbonate, et en présence d'ATP et de Mg<sup>2+</sup> (FIGURE 30). Cependant, aucune de ces études ne parvint à identifier la ou les enzyme(s) responsable(s) de cette synthèse, et le rendement de réaction *in vitro*, à partir d'extraits cellulaires, restait faible et variable. En raison de ces difficultés, on soupçonna donc l'intervention de plusieurs enzymes et le passage par des intermédiaires activés, comme du carbamoyl-phosphate. A défaut de réussir à isoler les enzymes de synthèse de t<sup>6</sup>A, plusieurs travaux apportèrent des informations concernant les modalités de fonctionnement de cette machinerie.

#### b) On ne fait pas de recherche sans casser des œufs : caractérisation de la machinerie de synthèse de t<sup>6</sup>A Eucaryote

En utilisant des extraits solubles cytoplasmiques de cellules de Xénope, et des ARNt extraits de cellules humaines, il fut montré que t<sup>6</sup>A n'est synthétisée sur l'ARNt qu'après la maturation des extrémités 3' et 5' (KOSKI AND CLARKSON, 1982, HARADA 1984). Mais la caractérisation de la machinerie de synthèse de t<sup>6</sup>A chez les Eucaryotes fit un bond en avant grâce à une approche originale: l'utilisation d'ovocytes de Xénope comme « tubes à essai cellulaires ». Au lieu d'étudier *in vitro* la machinerie de synthèse de t<sup>6</sup>A, l'idée était de la laisser opérer *in cellulo* dans ces ovocytes, en fournissant par micro-injection des substrats ARNt radiomarqués, que l'on récupérait ensuite par extraction (FIGURE 31). Ainsi, sans même connaître les enzymes responsables de cette modification, il était possible de suivre la cinétique de cette réaction et d'identifier les déterminants de la reconnaissance de l'ARNt substrat. Deux travaux s'appuyant sur cette approche étudièrent donc les modalités de synthèse de t<sup>6</sup>A (GROSJEAN 1987, MORIN 1998).

L'injection de variants d'ARNt de Levure, mutés à différentes positions pour altérer la séquence ou la structure tridimensionnelle de la molécule, permit de mettre en évidence deux paramètres critiques pour la synthèse de t<sup>6</sup>A chez le Xénope: d'une part, la

Bras D et T = structure  
2D (trèfle) et 3D (L)



**FIGURE 32 :** Représentation des déterminants moléculaires nécessaires à la machinerie de biosynthèse de t<sup>6</sup>A chez *Xenopus laevis*, déterminé par microinjection d'ARNt *in cellulo*.

Les déterminants strictement nécessaires pour la synthèse de t<sup>6</sup>A sont représentés en rouge. La présence des bras D et T, indispensables à la structuration tridimensionnelle de l'ARNt, témoigne d'une enzyme de classe II.

Au niveau de la boucle anticodon, le motif U36-A37 est strictement requis pour la synthèse de t<sup>6</sup>A (rouge). Un A en position 38 favorise fortement la synthèse de t<sup>6</sup>A (en orange). La présence d'un C32 et d'un U33 sont généralement associés à la présence de t<sup>6</sup>A (en jaune). d'après Morin 1998

composition de la boucle anticodon, et d'autre part la structure tertiaire de l'ARNt. Ainsi, l'A<sub>37</sub> de l'ARNt<sup>Met</sup> initiateur de *S. cerevisiae* (anticodon CAU, naturellement porteur de t<sup>6</sup>A chez cet organisme) est modifié en t<sup>6</sup>A très rapidement dans les ovocytes de Xénope (100% de modification en quelques heures). Les différents mutants de cet ARNt montrent que cette modification requiert d'une manière stricte la présence du couple U<sub>36</sub>-A<sub>37</sub>. La présence d'un A en position 38 augmente fortement cette synthèse. Cette observation est corrélée par le fait que quasiment tous les ARNt séquencés (issus de différents organismes) ayant un t<sup>6</sup>A<sub>37</sub> possèdent ce triplet de nucléotides : U<sub>36</sub>A<sub>37</sub>A<sub>38</sub>, qui semble donc être un déterminant universel de reconnaissance par les enzymes de biosynthèse de t<sup>6</sup>A (FIGURE 32).

De plus, un ARNt<sup>Val</sup> de Levure (anticodon UAC, qui ne porte pas naturellement t<sup>6</sup>A) dont la boucle anticodon est mutée pour être identique à la boucle anticodon de l'ARNt<sup>Met</sup> devient un excellent substrat de la réaction. En comparant cet ARNt muté avec les substrats ARNt<sup>Thr</sup> et ARNt<sup>Ile</sup>, on en arrive à une séquence consensus qui constitue la meilleure identité pour la reconnaissance par la machinerie t<sup>6</sup>A du Xénope : C<sub>32</sub>U<sub>33</sub>N<sub>34</sub>N<sub>35</sub>**U<sub>36</sub>A<sub>37</sub>A<sub>38</sub>**. Cependant, ce consensus ne peut être généralisable, car certains organismes n'ont pas de C en position 32 (*E. coli*). La taille de la boucle anticodon quant à elle n'influence pas cette réaction, tout comme le type de nucléoside en position 34 et 35, ou la composition de la tige anticodon. Ainsi, malgré la présence stricte du motif **U<sub>36</sub>A<sub>37</sub>A<sub>38</sub>**, une certaine variabilité semble tolérée dans le reste de la boucle pour la reconnaissance du substrat, permettant d'expliquer la présence de t<sup>6</sup>A de manière universelle malgré des séquences d'ARNt variables d'un organisme à l'autre.

Au niveau structural, des perturbations locales de certaines interactions clés, qui n'affectent pas la structure générale de l'ARNt, ont peu d'impact sur la synthèse de t<sup>6</sup>A. En revanche, des perturbations drastiques de la structure en feuille de trèfle, comme la délétion des bras D et T, abolissent totalement la synthèse de t<sup>6</sup>A. La machinerie eucaryote fait donc partie des enzymes des modifications de classe II, qui requiert strictement une conformation tridimensionnelle correcte.

Lors de ces expériences, plusieurs autres modifications (D<sub>47</sub>, m<sup>5</sup>C<sub>48</sub>, m<sup>1</sup>A<sub>58</sub>) sont synthétisées de manière concomitante à t<sup>6</sup>A, ce qui ne permet pas d'exclure que certaines modifications préalables soient requises ou favorisent la synthèse de t<sup>6</sup>A.

	Eukaryotes	Archées	Bactéries (et organites)
Codon d'initiation	AUG	AUG (GUG, UUG)	AUG (GUG, UUG)
ARNt Met initiateur	t <sup>6</sup> A (37)	pas de t <sup>6</sup> A	pas de t <sup>6</sup> A
ARNt Met élongateur	t <sup>6</sup> A (37)	t <sup>6</sup> A (37)	t <sup>6</sup> A (37)

ARNt<sup>Met</sup> initiateur de *E.coli*  
 injecté dans ovocyte de Xénope  
 (Eucaryote)  
 =  
 A<sub>37</sub> modifié en t<sup>6</sup>A  
 (Grosjean et al 1987)

**FIGURE 33 :** L'ARNt<sup>Met</sup> initiateur révèle des différences entre les machineries de synthèse t<sup>6</sup>A chez les trois domaines du vivant



### c) Des machineries différentes selon les domaines du vivant ?

Chez les Bactéries, l'ARNt<sup>fMet</sup> initiateur ne porte naturellement pas t<sup>6</sup>A, bien qu'il possède les éléments structuraux décrits précédemment. Or, de manière surprenante, l'A<sub>37</sub> de l'ARNt<sup>fMet</sup> (*E. coli*) est modifié très rapidement en t<sup>6</sup>A dans les cellules de Xénope ! (GROSJEAN ET AL., 1987). Ainsi, la machinerie eucaryote peut reconnaître cet ARNt comme substrat. Cela implique que l'absence de t<sup>6</sup>A *in vivo* au niveau de l'ARNt<sup>Met</sup> chez *E. coli* ne repose pas sur une particularité intrinsèque à cet ARNt (comme par exemple une richesse particulière en G/C), mais plutôt sur des différences propres au système enzymatique bactérien, comme des modalités de reconnaissance différentes du système eucaryote ou la présence de facteurs de sélectivité spécifiques. (FIGURE 33). L'ARNt<sup>Met</sup> initiateur des Archées étant également dépourvu de t<sup>6</sup>A, une telle différence pourrait également exister concernant la machinerie de synthèse chez ces organismes, bien qu'aucune donnée ne permette de le confirmer pour le moment. Enfin, d'une manière intéressante, l'ARNt<sup>Ser (GCU)</sup> de la mitochondrie de bœuf ne possède pas de bras D (indispensable pour la synthèse de t<sup>6</sup>A au niveau cytoplasmique), mais porte t<sup>6</sup>A (ARCARI ET BROWNLEE, 1980), ce qui peut permettre de penser que la machinerie mitochondriale a également des éléments de reconnaissance qui diffèrent de la machinerie cytoplasmique.

Ainsi, l'injection d'ARNt dans des ovocytes est une approche qui compense l'absence d'enzymes purifiées disponibles. Elle a permis de déterminer précisément les éléments structuraux essentiels requis pour la machinerie t<sup>6</sup>A des eucaryotes : un motif U<sub>36</sub>A<sub>37</sub>A<sub>38</sub>, et une structure 2D/3D intacte. Cette approche a également permis de pointer des différences probables entre les machineries issues des trois domaines du vivant au niveau de leurs modalités de reconnaissance. Cependant, cette approche présente des limites : l'ovocyte de Xénope est une cellule « particulière » au niveau de son métabolisme, de multiples enzymes de modifications agissent simultanément sur le substrat injecté, et la microinjection est difficilement transposable à d'autres organismes, voire impossible. Elle ne peut donc pas remplacer les données issues d'études *in vitro* qui seraient réalisées avec des enzymes purifiées.



#### **d) La biosynthèse du dérivé de $ms^2t^6A$ : quand modification d'ARNt rime avec santé**

Alors que les enzymes de synthèse de  $t^6A$  restaient toujours mystérieuses, les chercheurs eurent plus de chance concernant la synthèse des dérivés  $m^6t^6A$ ,  $ms^2t^6A$ , et  $ct^6A$ . Le cas de  $ms^2t^6A$  est un cas particulièrement intéressant, car il illustre l'intérêt de la caractérisation des systèmes enzymatiques de biosynthèse de ces nucléosides modifiés.

Tout comme  $t^6A$ , on retrouve  $ms^2t^6A$  sur certains ARNt dans les trois domaines du vivant, mais sur un nombre très limité d'ARNt. Par exemple, chez l'Homme comme chez *Bacillus subtilis*, on le retrouve uniquement au niveau de l'ARNt<sup>Lys(UUU)</sup>. L'enzyme synthétisant ce dérivé a été découverte récemment chez *Bacillus subtilis* : il s'agit d'une méthylthiotransférase, appelé MtaB (pour methylthio-threonylcarbamoyl-adenosine transferase B), catalysant l'addition d'un groupement thio méthyl sur le groupement préalable  $t^6A$  (ANTON ET AL., 2010). Les Eucaryotes et les Archées présentent un orthologue de cette enzyme, nommé e-MtaB. (ARRAGAIN ET AL., 2010). Chez l'Homme et la souris, l'orthologue de cette enzyme est codé par le gène *cdkal-1* (pour « cdk5 regulatory associated proteine 1-like »). Or, depuis plusieurs années, on savait sans en connaître l'origine que des variations spécifiques dans la séquence du gène *cdkal1* étaient associées au développement du diabète de type 2, une maladie multifactorielle touchant près de 21 millions de personnes en Europe se traduisant par un déficit d'insuline. Finalement, après avoir découvert le rôle des orthologues de CDKAL1 chez les Bactéries, le lien entre modification d'ARNt et diabète a été démontré très récemment, mettant en jeu une cascade d'évènements surprenante (WEI ET AL., 2011). Des mutations dans le gène *ckal1* engendrent une protéine non fonctionnelle, qui conduit à l'absence de  $ms^2t^6A$  sur l'ARNt<sup>Lys(UUU)</sup>. Cela a pour conséquence des erreurs fréquentes dans la lecture du codon AAA et AAG, ainsi que des décalages de cadre de lecture. Or, le gène de l'insuline contient deux codons Lysine, et l'une des lysines est située stratégiquement au niveau du site de clivage entre les chaînes A et C de la pro-insuline. Une mauvaise lecture de ce codon provoquerait ainsi la production d'une pro-insuline impossible à maturer, d'où un déficit global d'insuline fonctionnelle et le développement du diabète de type II.

		<b>BACTERIE</b>		<b>EUCARYOTE</b>		<b>ARCHEE</b>
		<i>Escherichia coli</i>	<i>Bacillus subtilis</i>	<i>Saccharomyces cerevisiae</i>	<i>Homo Sapiens</i>	<i>Pyrococcus abyssi</i>
Universelles		<b>YrdC</b> RimN/TsaC	YrdC	<b>Sua5</b>	IRIP	<b>Sua5</b> PAB1302
		<b>YgjD</b> Gcp/TsaD	YdiE	<b>Kae1</b>	OSGEP	<b>Kae1</b> PAB1159
				[Qri7] <sub>mt</sub>	[OSGEPL] <sub>mt</sub>	
Eucaryote/Archée spécifiques				<b>Bud32</b>	PRPK	<b>Bud32</b> PAB1047
				<b>Cgi121</b>	Cgi121	<b>Cgi121</b> PAB1522
				<b>Pcc1</b>	NY-ESO (1,2,3)	<b>Pcc1</b> PAB3073
				Gon7/Pcc2		
Bactérie spécifiques		<b>YeaZ</b> TsaB	YdiC			
		<b>YjeE</b> TsaE	YdiB			

} **KEOPS**

**FIGURE 34 :** Tableau de nomenclature des différentes protéines citées et étudiées dans ce manuscrit, et impliquées dans la voie de biosynthèse de t<sup>6</sup>A dans les trois domaine du vivant. Les noms en gras sont ceux utilisés dans le texte.

Ainsi, la compréhension des voies de biosynthèse des nucléosides modifiés, en particulier de t<sup>6</sup>A, qui est présente chez l'Homme à la fois au niveau cytoplasmique et mitochondrial, est un enjeu majeur qui peut déboucher sur des applications et des découvertes inattendues. Si les méthodes classiques de biochimie *in vitro* et *in vivo* ne permettent pas d'identifier cette fameuse t<sup>6</sup>A synthétase, c'est vers des techniques *in silico* qu'il faut se tourner. L'idée est la suivante : t<sup>6</sup>A étant une modification universelle, peut-on identifier une famille de protéines universelles de fonction indéterminée, en tant que candidat à la t<sup>6</sup>A synthétase?

Ce raisonnement fut suivi en 2009 par l'équipe de Valérie de Crécy-Lagard (EL YACOUBI ET AL, 2009) qui identifia l'une des protéines intervenant dans la voie de biosynthèse de t<sup>6</sup>A : la protéine YrdC chez *E. coli* et son homologue Sua5 chez *S. cerevisiae*, membres du COG0009. Avant qu'ils ne soient impliqués dans la biosynthèse de t<sup>6</sup>A, les membres de ce COG ont été, depuis leur identification, étudiés de manière extensive et sont devenus aussi bien une source d'information que de questionnement. Nous allons donc revenir dans un premier temps sur les données issues de ces différents travaux.

NB : Dans la suite du texte, de nombreuses autres protéines vont être abordées en lien avec la synthèse de t<sup>6</sup>A. Ces protéines et leurs orthologues portent des noms différents selon les organismes d'origine ou les travaux au cours desquels ils ont été caractérisés. La **FIGURE 34** est un tableau récapitulatif de la nomenclature retrouvée dans la littérature. Les noms en gras seront ceux utilisés dans ce manuscrit. Cette nomenclature est discutée page 285.



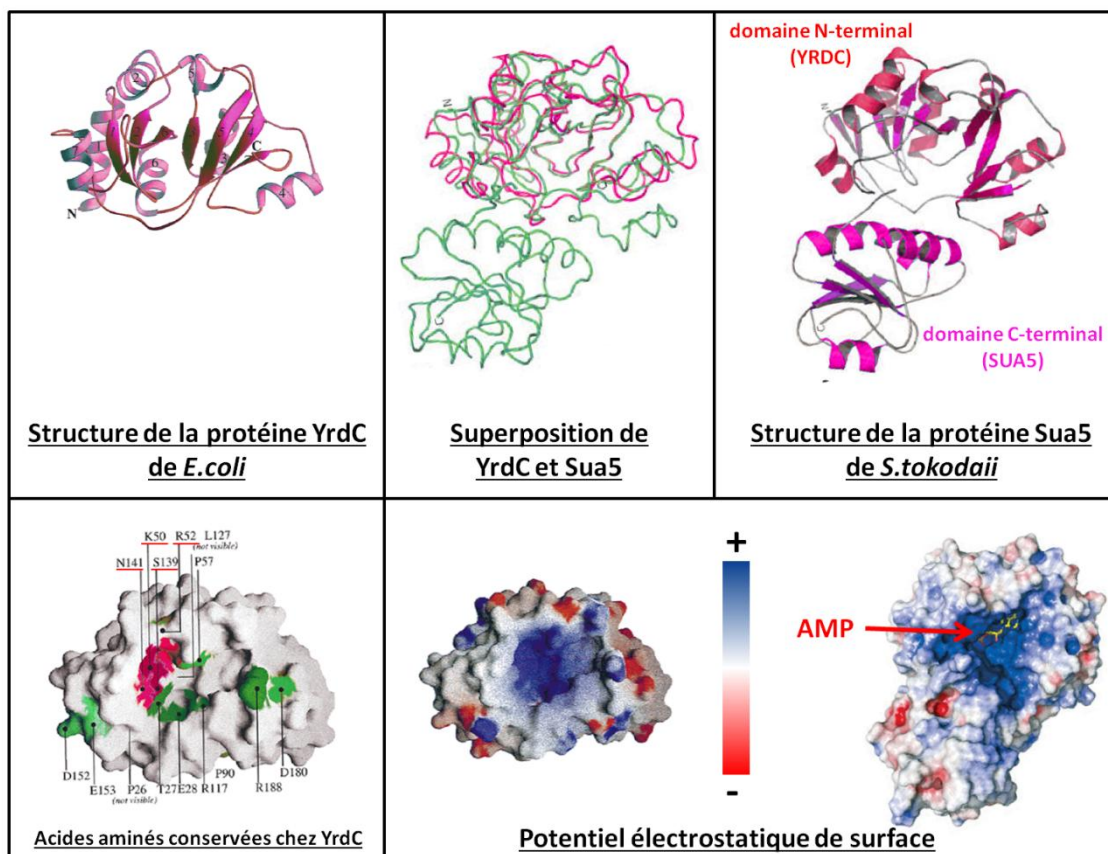
# Chapitre III : La Quête de la t<sup>6</sup>A synthase

## A - Sua5/YrdC (COG0009) : première pièce du puzzle de la t<sup>6</sup>A synthase

### 1 - Une protéine impliquée dans la traduction ?

En 1991 un ensemble de gènes furent identifiés chez *Saccharomyces cerevisiae* comme étant des supprimeurs d'un défaut d'initiation de la traduction causé par une mutation dans le gène *cyc1*, codant pour la protéine Cyc1 (le cytochrome C) (HAMPSEY ET AL., 1991). Dans une souche portant l'allèle *cyc1-1019*, une substitution conduisant à l'apparition d'un AUG aberrant en amont du cadre de lecture provoque une production de 2% de la quantité normale de la protéine. Huit gènes furent identifiés comme supprimeurs de cet effet et appelés *sua1-8* (pour *suppressor of upstream ATG*). Ainsi, lorsque l'un de ces gènes est inactivé, l'initiation sur cet AUG aberrant est réduite. Le gène *sua5*, codant une protéine de fonction inconnue, fut alors l'objet d'une étude plus poussée (NA ET AL., 1992). Ces études montrèrent que l'effet de *sua5* ne se situe probablement pas à un niveau transcriptionnel. De plus, la délétion de *sua5* a pour conséquence des effets pléiotropiques (croissance très ralentie, impossibilité de croître en présence stricte de glycérol et de lactate, indiquant un défaut mitochondrial, et une diminution du niveau de cytochrome b), impliquant ce gène dans des processus cellulaires fondamentaux.

Chez l'Homme, l'orthologue de Sua5, appelé YrdC<sub>H</sub>, fut identifié en 2003 par la recherche de partenaires de la protéine RBBP10 (*retinoblastoma binding protein 10*, un partenaire du suppresseur de tumeur Rb1). L'expression ubiquitaire de YrdC<sub>H</sub> dans les tissus humains, et particulièrement dans les tissus ayant un fort taux de synthèse protéique (foie, cerveau et tumeurs) suggéra un rôle au niveau de la traduction (CHEN ET AL. 2003). Une autre étude parallèle chez la Souris, recherchant les gènes activés par des conditions d'ischémie (c'est-à-dire une carence d'oxygène), identifie IRIP (*ischemia/reperfusion inducible protein*) comme l'homologue murin de la famille YrdC. Comme chez l'Homme, le gène codant cette protéine



**FIGURE 35 :** Comparaison des structures des protéines YrdC d'*Escherichia coli* de Sua5 de *Sulfolobus tokodaii*.

Le cadre supérieur gauche montre la structure de la protéine YrdC d'*Escherichia coli*, et le cadre supérieur droit celle de Sua5 de *Sulfolobus tokodaii*. Les domaines YRDC (N-terminal) et SUA5 (C-terminal) sont indiqués, et le « linker » de 20 résidus entre les deux domaines n'apparaît pas dans cette structure. Le cadran supérieur central montre la superposition des deux structures. Dans le cadre inférieur gauche, les acides aminés conservés chez les orthologues de YrdC sont colorés en vert, et les acides aminés conservés de la tétrade catalytique (KxxR SxxN) sont colorés en rouge.

Le cadran inférieur droit montre le potentiel électrostatique de surface des deux protéines, où apparaît une cavité fortement électropositive (en bleu). La molécule d'AMP présente dans le site actif de Sua5 est indiquée. D'après Teplova *et al.*, 2000, Arani *et al.*, 2008

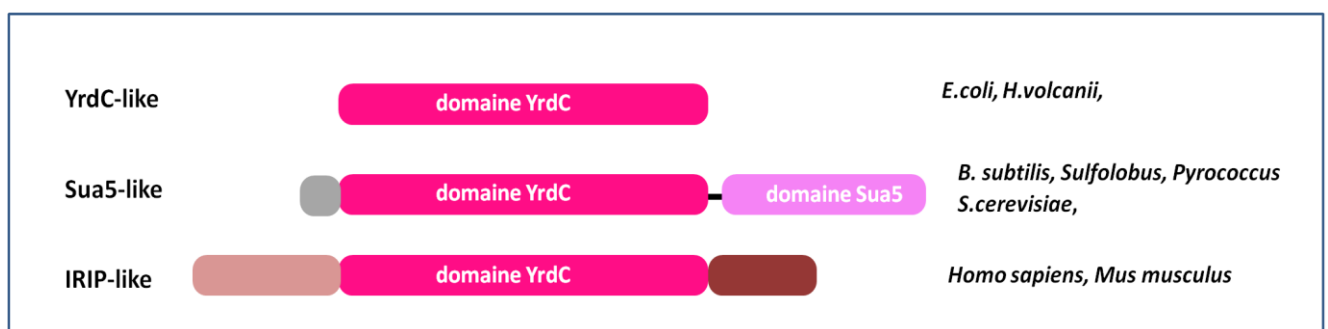


est exprimé de façon ubiquitaire. Cette étude, menée sur des lignées cellulaires par inactivation ou surexpression du gène, suggéra que IRIP/YrdC<sub>H</sub> joue un rôle dans la régulation de divers transporteurs membranaires (JIANG ET AL., 2005)

Chez la bactérie *Escherichia coli* enfin, la fonction de l'homologue YrdC a été également étudiée (KACZANOWSKA ET AL., 2005). Une souche portant une mutation dans le gène *prfA*, codant l'un des facteurs de terminaison de la traduction, présente un phénotype sensible à la température. Il a été montré qu'une délétion du gène *yrdC* supprime ce phénotype. La délétion de *yrdC* se révèle être inséparable de la délétion de *prfA1*, ce qui suggère que YrdC est une protéine essentielle dans les conditions normales de culture en laboratoire. De manière intéressante, ce double mutant présente une augmentation de la quantité d'ARNr 17S, un précurseur immature de l'ARNr 16S. Ainsi, il a été suggéré que YrdC pourrait intervenir dans la régulation de la traduction via la maturation du ribosome, bien qu'aucune interaction directe entre YrdC et un ARNr n'ait été démontrée. Le nom de RimN (ribose maturation) fut alors proposé pour remplacer YrdC.

Plus récemment, une nouvelle étude sur la protéine Sua5 de Levure vint ajouter un doute concernant la fonction de cette protéine (MENG ET AL., 2010). Cette étude révèle que la délétion de Sua5 est associée à une réduction progressive de la taille des télomères, et des expériences de retard sur gel *in vitro* montrent une fixation spécifique de cette protéine sur de l'ADN télomérique. Cela suggère que cette protéine pourrait intervenir dans l'homéostasie des télomères par interaction directe.

Ainsi, ces études *in vivo*, établies chez différents organismes et selon différentes techniques, indiquent que Sua5/YrdC doit jouer un rôle fondamental au niveau cellulaire, peut-être lié à la traduction, sans toutefois fournir un modèle satisfaisant et cohérent quant à sa fonction biologique.



**FIGURE 36 :** Représentation des orthologues de YrdC/Sua5 représentatifs de différents groupes d'organismes.

Les trois types d'orthologues sont représentés. Le type « YrdC-like », est retrouvé chez certaines Archées et Bactéries, dont *E. coli* et *H. volcanii*. Le type « Sua5 like » est composé de la fusion d'un domaine YrdC fusionné à un domaine SUA5 en position C-terminale, et est retrouvé chez certaines Archées (*P. abyssi*, *S. tokodaii*) Bactéries (*B. subtilis*), Eucaryotes (*S. cerevisiae*).

Le type « IRIP-like » est composé du domaine YrdC, fusionné à des polypeptides en position N et C terminales dont la structure n'est pas connue. On le retrouve chez l'Homme et les Mammifères.

## 2 - Une famille universelle... mais hétérogène :

Chez tous les organismes ayant fait l'objet des études précédemment décrites, les gènes appelés Sua5 (chez *S. cerevisiae*) ou YrdC (chez *H.sapiens* et *E. coli*) ont comme point commun la présence d'un domaine conservé appelé YRDC. Le caractère universel de cette famille, associé à une lacune concernant sa fonction cellulaire, a placé les membres de ce COG dans le Top 10 des « cibles prioritaires d'étude » en 2004 (GALPERIN AND KOONIN., 2004). Malgré cette universalité, il apparaît qu'au niveau structural et biochimique, une certaine hétérogénéité caractérise cette famille.

Chez *Escherichia coli*, la structure de la protéine YrdC, qui n'est constituée que de ce domaine conservé éponyme, a été résolue (FIGURE 35). Ce domaine présente une large cavité positive, pouvant constituer une zone d'interaction avec un acide nucléique. *In vitro*, l'utilisation de la technique d'extinction de la fluorescence des tryptophanes (en anglais *fluorescence quenching*) a montré que cette protéine présente en effet une fixation à l'ARN double brin et à l'ARNt dont la constante d'affinité est de l'ordre de 0,2-0,7 $\mu$ M. (TEPLOVA ET AL., 2000). La structure de la protéine Sua5 de l'Archée *Sulfolobus tokodaii* a été également résolue une première fois en 2008 (AGARI ET AL, 2008) (FIGURE 35). Comme chez son homologue bactérien, cette protéine présente le domaine conservé YRDC (en position N-terminale), mais possède également un domaine supplémentaire C-terminal nommé SUA5. Tout comme chez *E. coli*, le domaine YRDC présente une cavité électropositive. En revanche, au sein de cette dépression, une molécule d'AMP est co-cristallisée, interagissant avec une tétrade d'acides aminés (KxR-SN) conservée chez tous les orthologues de cette famille, y compris YrdC de *E. coli* (TEPLOVA ET AL., 2000, EL YACOUBI ET AL.2009). La présence de cet AMP est corrélée avec l'activité *in vitro* de cette protéine, qui hydrolyse l'ATP en AMP. Plus récemment, une nouvelle structure a été obtenue, avec au sein de la cavité une molécule d'AMPPNP et de L-thréonine (AGARI ET AL. 2008, nous en reparlerons par la suite). Le domaine C-terminal SUA5, quant à lui, montre une similitude structurale avec certaines GTPases.



Au bilan, d'après les banques de données protéiques, il apparaît que les Archées et les Bactéries présentent un orthologue de YrdC/Sua5 (parfois en plusieurs exemplaires) sous l'une de ces deux configurations : soit sous forme d'un domaine YRDC indépendant (comme chez *E. coli* et *H. volcanii*), soit sous forme d'une fusion YRDC-SUA5 (comme chez *S. tokodaii* ou *B. subtilis*). (FIGURE 36) Pour le moment, aucune étude exhaustive de la distribution phylogénétique de ces deux différentes formes n'a été effectuée.

En ce qui concerne les Eucaryotes, les Champignons, comprenant la Levure *S. cerevisiae*, présentent une protéine ayant l'organisation « Sua5 », avec un domaine YRDC fusionné à un domaine SUA5. De manière intéressante, les deux domaines séparés semblent ne pas présenter les propriétés de fixation aux acides nucléiques que présente la protéine entière (MENG ET AL., 2010). Chez l'Homme et les autres mammifères, la structure de l'orthologue est très différente : on retrouve le domaine conservé universel YrdC, au sein d'un polypeptide présentant d'autres domaines non caractérisés (FIGURE 34), mais présentant des motifs identifiables d'après Chan et al : un motif « *GTP-elongateur factor* », présent chez les facteurs d'élongation universel Ef-1 et Ef-2 et caractéristique des protéines fixant les ARNt non initiateurs au ribosome, ainsi qu'un motif « *leucine zipper* », impliqué dans la dimérisation et la fixation aux acides nucléiques chez d'autres protéines. (CHEN ET AL., 2003)

Au bilan, les protéines de cette famille se caractérisent par la présence systématique d'un domaine YrdC conservé, qui présente une cavité électro-positive qui pourrait être associée à une certaine affinité pour les acides nucléiques ou les nucléotides. Mais tout comme les données issues d'expériences de génétique, les données biochimiques et structurales sont hétérogènes et ne permettent pas d'établir un modèle cohérent pour expliquer la fonction de cette protéine.



### **Un paralogue d'YrdC :**

Bien que la plupart des organismes ne présentent qu'une version de Sua5/YrdC, certains organismes présentent un autre exemplaire de cette famille : la protéine YciO (d'après *Escherichia coli*). La structure de cette protéine (JIA ET AL., 2002) montre une certaine ressemblance avec YrdC, malgré des différences notables, principalement au niveau du site actif potentiel : la tétrade impliquée dans la fixation du nucléotide n'est pas conservée. Des expériences de complémentation génétique ont montré que ce paralogue ne semble pas remplir la même fonction que YrdC et représenterait une famille de protéines au rôle différent (EL YACOUBI ET AL., 2009, GERDES ET AL., 2011), dont nous ne parlerons pas par la suite.

### **3 - Un rôle dans la biosynthèse de t<sup>6</sup>A :**

**Les nombreuses données issues des ces différentes études soulèvent donc plus de questions qu'elles n'apportent de réponses. Cette famille de protéines présente-t-elle une fonction unique et conservée ? Quelle est cette fonction ? Comment expliquer les différentes caractéristiques observées *in vivo* et *in vitro* ? Si l'approche de génétique classique ne peut fournir une réponse satisfaisante, c'est de la génomique que viendra la solution.**

En 2009, le lien entre t<sup>6</sup>A et la famille YrdC/Sua5 va ainsi être établi. A la recherche de l'enzyme responsable de la synthèse de t<sup>6</sup>A, *El Yacoubi et al.* s'intéressent alors à cette famille pour 4 raisons :

- C'est une famille universelle, présentant un orthologue chez tous les organismes séquencés, ce qui correspond à la distribution théorique de t<sup>6</sup>A.
- Sa fonction reste indéterminée, mais il semble y avoir un lien avec la traduction.
- Cette protéine semble avoir une affinité pour les acides nucléiques, dont les ARNt.
- Enfin, on retrouve le domaine YrdC par similarité de séquence au niveau de plusieurs protéines à activité carbamoyl-transférase (HypF, NodU) ce qui correspond à l'activité prévue pour la synthèse de t<sup>6</sup>A (ce dernier point sera détaillé par la suite).



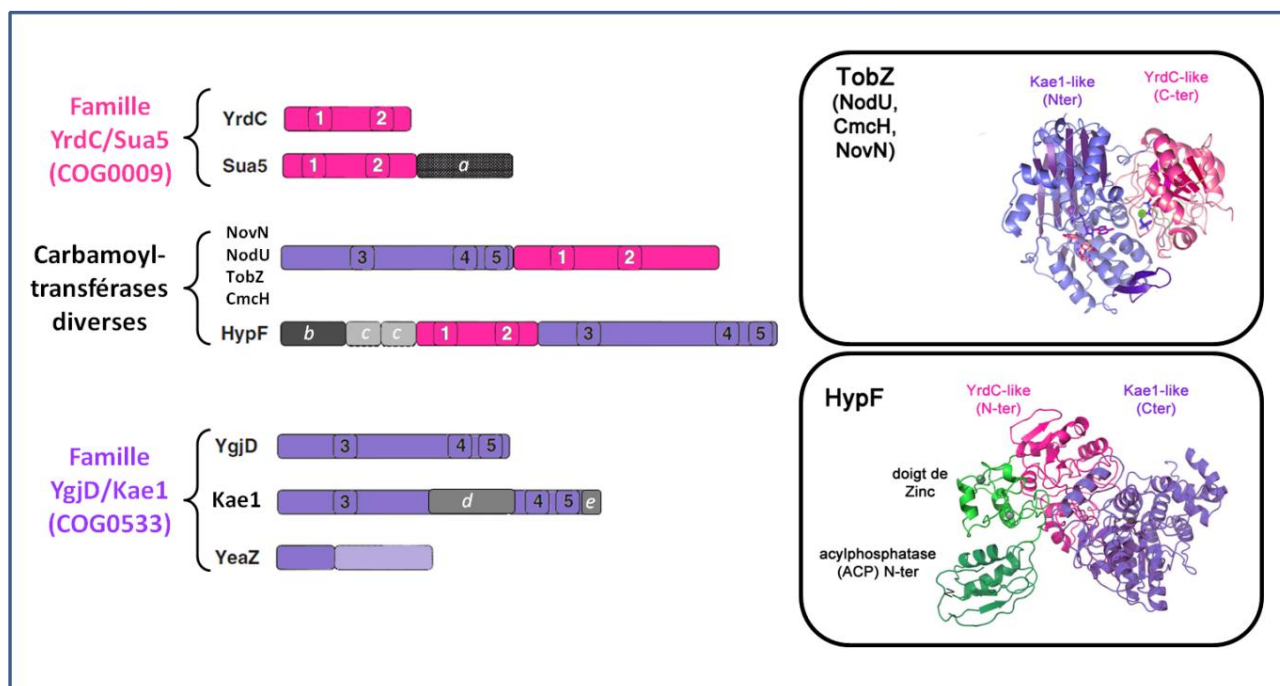


En se basant sur ces hypothèses, il fut alors montré que la délétion de Sua5 chez *S. cerevisiae* est effectivement associée à une absence de t<sup>6</sup>A au niveau des ARNt. L'expression des gènes orthologues de Sua5 issus des bactéries *E. coli*, *B. subtilis* et de l'archée *Methanococcus maripaludis* rétablissent la présence de t<sup>6</sup>A chez ce mutant. Inversement, une souche d'*E. coli* délétée pour YrdC peut être sauvée par l'expression de ces orthologues. Ces complémentations indiquent que la fonction biologique de la famille Sua5/YrdC est conservée au sein du monde vivant.

Cependant, alors que la délétion de Sua5 chez *S. cerevisiae* n'est pas létale mais affecte fortement la croissance cellulaire, la délétion de YrdC chez *E. coli* ne conduit pas à un organisme viable. Cela suggère des différences quant à l'importance de l'activité de cette protéine chez ces deux organismes, traduisant peut-être une importance différente de t<sup>6</sup>A entre *S. cerevisiae* et *E. coli*.

Enfin, le rôle direct de YrdC/Sua5 dans la biosynthèse de t<sup>6</sup>A fut confortée par la mise en évidence d'une forte affinité de l'orthologue de *E. coli* pour les ARNt et l'ATP (EL YACOUBI ET AL., 2009), mais aussi la thréonine (HARRIS ET AL., 2011), qui sont les précurseurs de t<sup>6</sup>A. De même, une molécule de L-thréonine et d'AMPPNP a été décrite dans la structure de Sua5 de *Sulfolobus tokodaii* (KURATANI ET AL., 2011). En revanche, aucune incorporation de thréonine au niveau des ARNt ne fut obtenue avec la protéine purifiée YrdC d'*E. coli*. Ainsi, il fut suggéré qu'un autre partenaire non identifié soit requis dans la synthèse de t<sup>6</sup>A.

**Les études précédentes n'ayant pas identifié de façon certaine des partenaires conservés à YrdC/Sua5, comment trouver cette enzyme manquante ?**



**FIGURE 37 : Comparaison des protéines contenant les domaines structuraux YrdC et YgjD.**

Chaque protéine est représentée comme une barre horizontale, dont la taille correspond approximativement au nombre de résidus. Le domaine YrdC-like est coloré en rose, le domaine YgjD-like en violet. a= domaine SUA5. b= domaine acyl-phosphatase. c= domaines en doigt de zinc. La position des motifs conservés dans les deux familles de domaines sont indiqués dans des encarts numérotés. 1 et 2 : motif conservé KRSN impliqué dans l'interaction avec le ligand. 3 : fixation du métal. 4 = fixation du nucléotide, 5= fixation du métal. La protéine YeaZ présente un domaine C-terminal divergent par rapport aux protéines YgjD et Kae1.

Dans les encarts de droite, les structures des carbamoyl-transférases TobZ (en haut) et HypF (en bas) sont indiquées. Les couleurs et les domaines correspondant aux figurés de gauche sont indiqués. d'après El Yacoubi *et al.*, 2011.

## 4 - Les carbamoyl-transférases HypF/NodU/CmcH/TobZ : la clé du puzzle

Les protéines HypF et NodU appartiennent à deux familles distinctes de carbamoyl-transférases. Ces protéines contiennent chacune deux domaines protéiques principaux. L'un des ces domaines présente une forte similarité de séquence avec la protéine YrdC, ce qui a servi de piste initiale pour l'implication de YrdC au niveau de la biosynthèse de t<sup>6</sup>A (EL YACOUBI ET AL., 2009). L'autre domaine, en revanche, ne trouve aucune correspondance dans les banques de données au niveau de sa séquence.

La résolution de la structure entière de HypF (PETKUN ET AL. 2011) confirma la similarité structurale du domaine intermédiaire avec YrdC, et montra de manière surprenante que le domaine C-terminal présente une forte similarité structurale avec une autre protéine universelle : la protéine Kae1 de *Pyrococcus abyssi*, dont il sera question dans la partie suivante (FIGURE 37). De même, la résolution plus récente de la structure de la protéine TobZ, proche de NodU (l'autre famille de carbamoyl-transférase) montre que cette protéine est également constituée de deux domaines structuraux YrdC et Kae1 fusionnés, mais dans un ordre inversé (PARTHIER ET AL., 2012) (FIGURE 37).

**Ainsi, de manière indépendante, des domaines structuralement proches de YrdC et de Kae1 sont retrouvés fusionnés au sein de différentes enzymes catalysant des réactions de carbamoylation, réaction chimique prévue pour la biosynthèse de t<sup>6</sup>A. Cette observation mena à une hypothèse qui fut confirmée en 2011: la protéine Kae1 interviendrait aux côtés de YrdC/Sua5 dans la voie de biosynthèse de t<sup>6</sup>A. Avant d'aborder ces travaux, nous allons revenir sur la caractérisation de Kae1 et de ses orthologues.**



## B - Kae1/YgjD (COG0533) : la seconde pièce du puzzle de la t<sup>6</sup>A synthétase

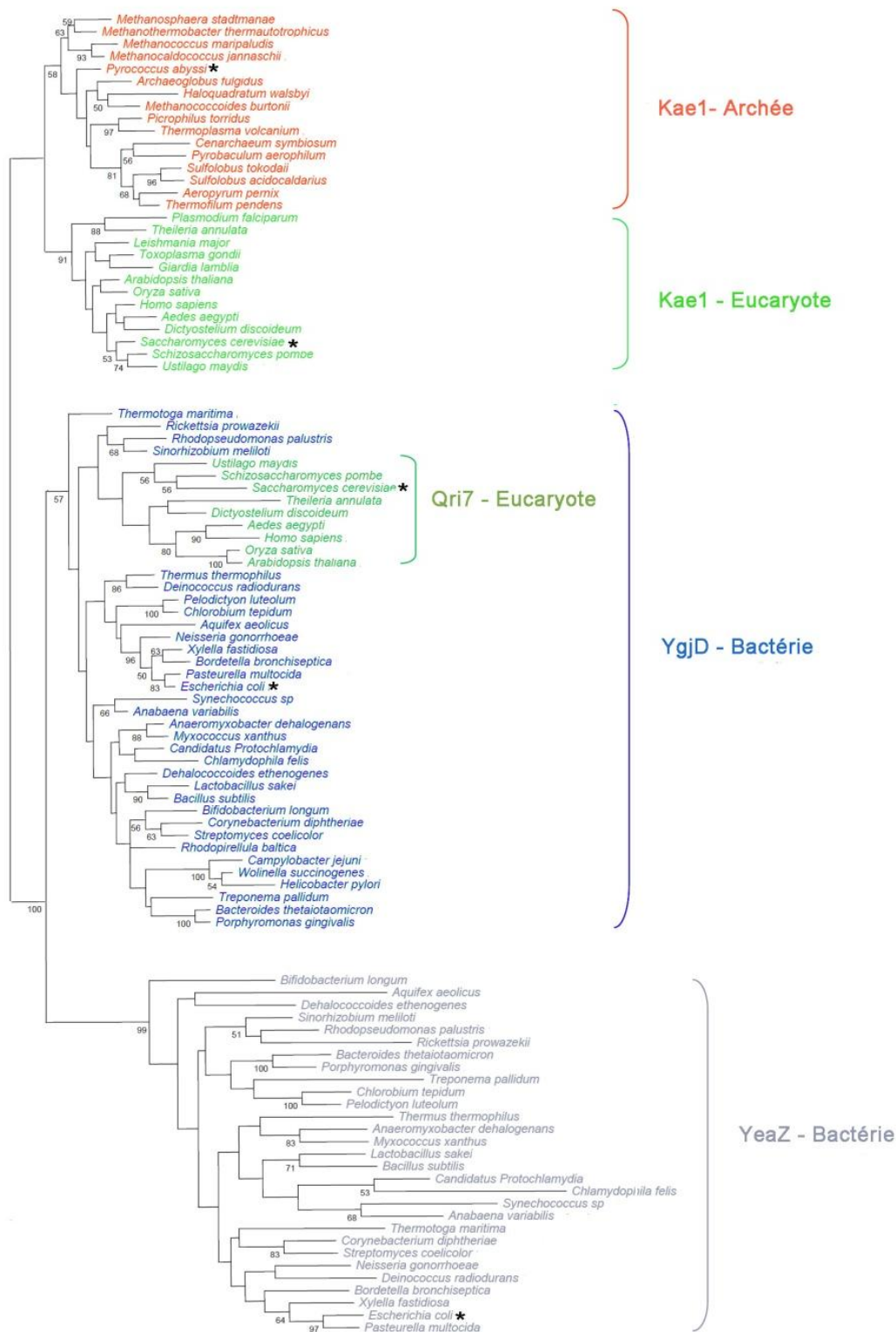
### 1 - YgjD/Kae1, une famille universelle

En 1987, le séquençage d'un fragment du génome d'*Escherichia coli* identifie un **nouveau cadre de lecture ouverte**, OrfX, appelé par la suite **ygjD** (NESSIN ET AL., 1987). La même année, une étude identifia ce gène comme étant l'orthologue d'un gène codant une O-glycosialoprotéase sécrétée par la bactérie *Pasteurella haemolytica*. Le gène *ygjD* fut ainsi annoté « putative metallo-protéase », et nommé **gcp** (ABUDLLAH ET AL. 1987)

L'une des premières études de génomique comparative, en 1996, basé sur la comparaison des génomes récemment séquencés de *Mycobacteria genitalium* et *Haemophilum influenzae*, identifia *ygjD* comme candidat susceptible de faire partie de « l'ensemble minimum de gènes requis pour une vie cellulaire indépendante », car conservé chez ces deux organismes. Cependant, il en fut exclu en raison de son annotation, l'activité glycoprotéase étant considérée comme spécifique de la vie parasitaire (KOONIN ET AL. 1996)

Finalement, deux ans plus tard, une nouvelle étude de génomique comparative suggéra que *ygjD* devrait à juste titre faire partie de cet ensemble, ce gène apparaissant comme essentiel chez les bactéries *E. coli* et *B. subtilis* (ARIGONI ET AL, 1998).

Grâce aux données issues du séquençage des génomes, il apparut rapidement que tous les organismes séquencés possèdent au moins un homologue du gène *ygjD* (KOONIN ET AL., 1999), confortant la place de cette famille, classée sous l'ensemble **COG0533**, au sein du « génome minimal » théorique (KOONIN ET AL., 2003). Tout comme la famille de Sua5/YrdC, elle fut placée dans le Top 10 des protéines conservées à étudier prioritairement (KOONIN ET AL., 2004), étant donné qu'aucune fonction précise ne lui n'était assignée.



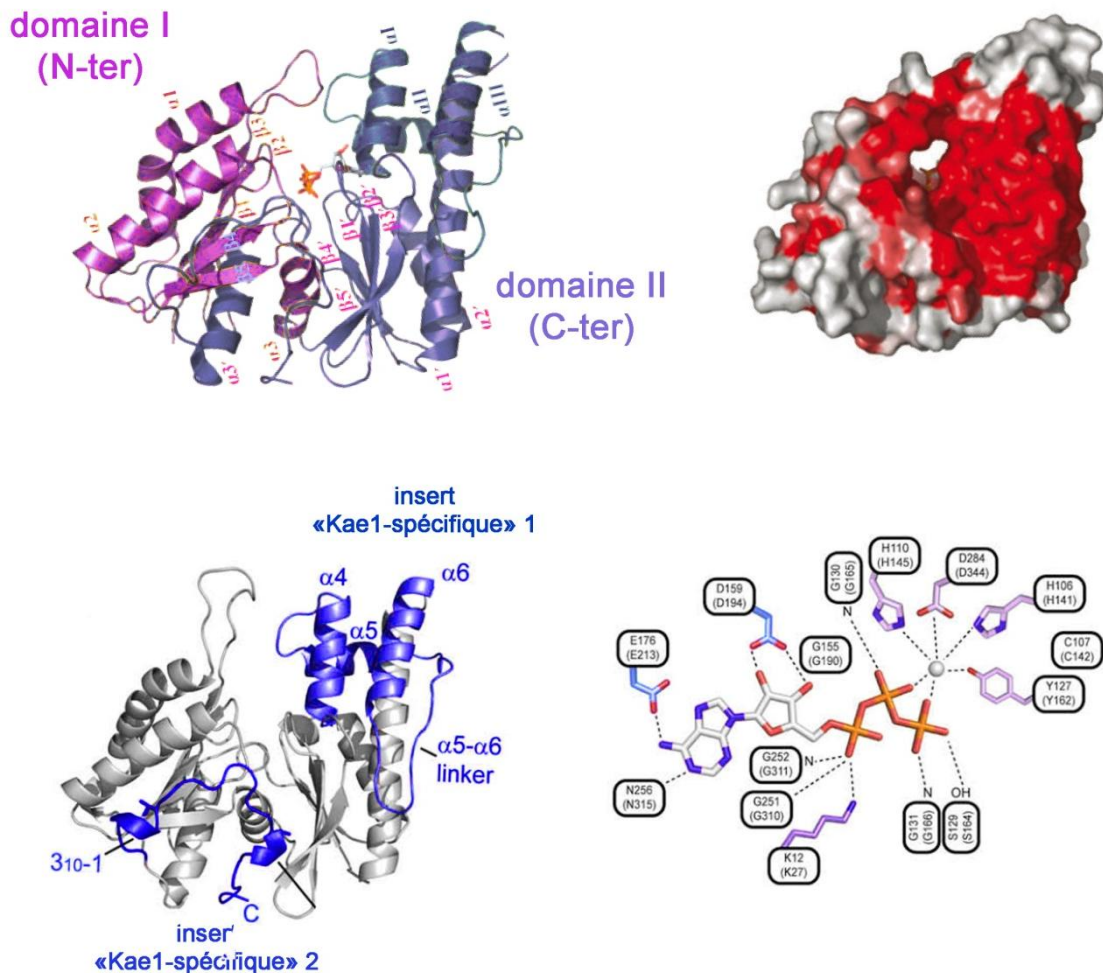
**FIGURE 38 : Phylogénie de la famille YgjD/Kae1**

747 orthologues ont été détectées dans les 409 génomes analysés, en un exemplaire minimum par génome, puis alignées et un arbre phylogénétique a été construit à partir de l'alignement obtenu. Seul un échantillon de 100 séquences issues d'espèces représentatives est montré. Les séquences d'Archées sont en rouge, les séquences de Bactéries en bleu, d'Eucaryotes en vert. Les séquences correspondant aux orthologues de YeaZ, un paralogue d'YgjD, sont colorées en gris. Les séquences Eucaryotes branchant au sein du groupe des Bactéries sont des orthologue de Qri7 de *S. cerevisiae*, et correspondent aux versions mitochondriales. d'après Hecker *et al.* 2007

L'étude phylogénétique de cette famille (HECKER ET AL., 2007) confirma son universalité: seuls les génomes extrêmement réduits des bactéries parasites *Carsonella rudeii* et *Sulcia mulleri* ne présentent pas d'homologues. Les Archées ne présentent qu'un unique orthologue de cette famille, alors que les Eucaryotes présentent deux homologues (appelés respectivement Kae1 et Qri7 chez *S. cerevisiae*), les Bactéries présentent également deux homologues (YgjD et YeaZ, respectivement). La topologie de l'arbre obtenu permet de comprendre cette répartition et montre une distribution divisée en 3 grands ensembles (FIGURE 38):

- Une première branche « *Kae1-like* », regroupe d'une part l'ensemble des orthologues d'Archées, et d'autre part les orthologues du gène *kae1* de *S. cerevisiae*, l'une des deux « versions » eucaryotes, considérée comme la version nucléaire/cytoplasmique.
- Une seconde branche regroupe les orthologues bactériens du gène *ygjD* de *E. coli*. De manière intéressante, les orthologues du gène *qri7* de *S. cerevisiae* (la seconde version eucaryote) sont regroupés au sein de cette branche bactérienne, plus précisément à proximité des gènes *ygjD* d' $\alpha$ -protéobactéries. Cette position phylogénétique traduit une origine mitochondriale, qui a été confirmée en 2009 par la mise en évidence de l'importation de Qri7 dans la mitochondrie chez *S. cerevisiae* et *C. elegans* (OBERTO ET AL., 2009, HAUSSUEHL ET AL., 2009).
- Une branche à part, proche de la branche « *YgjD* » bactérienne, regroupe tous les orthologues bactériens du gène *yeaZ* de *E. coli*. Ces gènes paralogues constituent une duplication de gène spécifique aux Bactéries, et sont considérés comme un COG distinct (COG1214) en raison du degré de divergence avec YgjD (chez *E. coli* par exemple YgjD et YeaZ ne partagent que 29% d'identité). Nous reviendrons dans le chapitre suivant sur la protéine YeaZ et son lien fonctionnel avec YgjD.

**Ainsi, la famille Kae1/YgjD est hautement conservée, présente chez tous les organismes y compris les organismes à génome réduit, et chez les mitochondries. Elle fait partie à juste titre du groupe de protéines probablement présentes chez le dernier ancêtre commun universel, LUCA. Cependant, la répartition phylogénétique montre une distinction entre une version « Bactérie/Mitochondrie » et une version « Archée/Eucaryote ».**



**FIGURE 39 : Structure de la protéine Kae1 de *Pyrococcus abyssi***

**Supérieur gauche** : mise en évidence de la structure de Kae1 de *P. abyssi*, avec ses deux domaines I et II. L'AMPPNP trouvé au niveau du site actif est représenté en orange. **Supérieur droit** : les résidus conservés chez la famille Kae1/YgjD sont colorés en rouge. La représentation en « surface » permet de visualiser la cavité du site actif, où se place l'AMPPNP. **Inférieur gauche** : les deux régions additionnelles, spécifiques à la famille Kae1/YgjD (par rapport à la super famille ASHKA) sont mises en évidence en bleu. **Inférieur droit** : visualisation du site catalytique de Kae1 de *M. jannaschii*. L'ATP a été modélisée d'après la molécule d'AMPPNP de Kae1 de *P. abyssi*, et l'atome de fer (en gris) est représenté. Les liaisons entre les résidus, le nucléotide et le fer sont indiquées en pointillés. La position des résidus sont indiqués dans les cadrans. Entre parenthèse sont indiqués les résidus conservés correspondant chez Kae1 de *S. cerevisiae*. d'après Hecker et al, 2007, Mao et al., 2008



## 2 - Propriétés structurales de la famille YgjD/Kae1 :

Une première étude prédictive structurale identifia chez la famille YgjD/Kae1 un repliement caractéristique semblable à différentes protéines ATPase, dont certaines possèdent une activité de chaperonne moléculaire : la chaperone universelle Hsp70, l'actine et son homologue bactérien MreB, et les kinases de sucres (KOONIN ET AL., 1999). De plus, une diade conservée d'histidines fut identifiée au sein de YgjD/Kae1, pouvant correspondre à un site de coordination d'un ion métallique, caractéristique des métallo-protéases.

Ces prédictions structurales ont été confirmées en 2007 par la résolution de la structure de Kae1 de *Pyrococcus abyssi* (HECKER ET AL., 2007). (FIGURE 39). Cette protéine présente effectivement un repliement qui la classe parmi la super famille ASKHA (*acetate/sugar kinase, Hsp70, actin*). Elle présente deux sous-domaines semblables, notés I et II. Deux régions particulières, appelées inserts spécifiques, l'un situé sur le domaine I et l'autre, relativement important, situé dans le domaine II, sont spécifiques à la famille Kae1/YgjD et n'existent pas sur les autres membres de la famille ASKHA. La co-cristallisation de Kae1 avec une molécule d'AMPPNP montre que ces deux domaines forment un tunnel conservé au sein duquel peut se fixer un nucléotide, sans provoquer de réarrangement structural de la protéine. Au sein de ce site est présent un atome de Fer qui intervient directement dans la fixation du nucléotide en interagissant avec le phosphate gamma. Ce mode de fixation de nucléotide est totalement unique et n'est retrouvé chez aucune autre protéine. Les acides aminés coordonnant le fer (une diade d'histidine, une tyrosine et un aspartate) sont hautement conservées chez les orthologues Archées et Eucaryotes (HECKER ET AL., 2007), et la mutation de ces résidus affecte fortement la croissance chez *S. cerevisiae*, ce qui suggère que ce métal doit jouer un rôle essentiel dans la fonction de la protéine (KISSELEVA 2006, MAO ET AL 2008).

La résolution récente de la structure de YgjD de *E. coli* (NICHOLS ET AL., 2013) montre que cette protéine présente un repliement très similaire à celui de Kae1, avec la présence d'un nucléotide (AMP ou ATPyS, selon le milieu de cristallisation utilisé) au sein du site actif. Cependant, chez les versions bactériennes et mitochondriales, seule la diade d'histidines est conservée au niveau du site actif, et la protéine YgjD présente un atome de zinc au lieu de l'atome de fer chez Kae1.



### 3 - Une protéine impliquée dans la stabilité et la dynamique du génome?

Initialement annotée comme métallo-protéase, aucune activité de ce type n'a été jamais été détectée chez aucun orthologue de cette famille (HECKER ET AL., 2007, MAO ET AL., 2008). Les auteurs des premiers travaux ont même suggéré une erreur d'identification lors des travaux initiaux (JIANG ET MELLORS, 2004). Au vue de l'universalité de cette famille, de très nombreuses études ont donc tenté de percer le mystère de la fonction biologique des protéines de la famille YgjD/Kae1 chez différents organismes modèles.\*

Il a été montré que Kae1 de *P. abyssi* se lie de manière aspécifique à l'ADN simple ou double brin d'une manière coopérative (HECKER ET AL., 2007), et forme des nucléofilaments. Lors de ces travaux, une activité endonucléase au niveau de sites apuriniques de l'ADN a également été détectée. De manière intéressante, chez l'Homme, l'homologue cytoplasmique de Kae1, nommé OSGEP (d'après l'annotation initiale O-sialoglyco-endopetidase protein), est un gène de ménage exprimé de façon ubiquitaire qui est co-transcrit avec un gène codant une endo-nucléase impliquée dans la réparation de l'ADN (SEKI ET AL., 2002). Chez *S. cerevisiae*, la délétion de Kae1 affecte fortement la croissance cellulaire, et cette protéine a été associée, au sein d'un complexe multoprotéique, à un rôle dans l'homéostasie des télomères d'une part (DOWNEY ET AL., 2006), et dans la régulation de la transcription de gènes essentiels d'autre part (KISSELEVA ET AL., 2006).

Chez les bactéries *E. coli* et chez *B. subtilis*, le gène *yjgD* est essentiel, et les mutants de délétion ne sont pas viables (ARIGONI ET AL., 1998), mais une déplétion progressive de YgjD chez *E. coli* conduit à plusieurs phénotypes pléiotropiques : une répartition inhabituelle de l'ADN, une perte du nucléoïde (OBERTO ET AL., 2009), ainsi qu'à un changement de la morphologie (HANDFORD ET AL., 2009). L'orthologue de YgjD chez *Deinococcus radiodurans*, qui n'est en revanche pas essentiel, et a été impliqué dans la réparation des *cross-links* d'ADN (ONODERA ET AL., 2013).

---

\* C'est d'ailleurs cette universalité qui est à l'origine de notre intérêt initial pour la protéine Kae1 (cf Préface)



Enfin, la délétion de la version mitochondriale (orthologue de Qri7) chez la plante *A.thaliana* conduit à des graines anormales et un arrêt très précoce du développement (HAUSSUEHL ET AL., 2009) et chez *S. cerevisiae* et *C. elegans*, l'absence de Qri7 a pour conséquence une fragmentation du génome mitochondrial ainsi qu'à une morphologie aberrante de cet organite, phénotypes similaires à ceux rencontrés chez *E. coli* (OBERTO ET AL., 2009)

L'ensemble de ces données indique que les membres de la famille YgjD/Kae1/Qri7 jouent un rôle biologique fondamental. Différents indices suggèrent que ces protéines pourraient être impliquées plus ou moins directement dans le maintien et la dynamique du génome, et il a été ainsi proposé pendant un temps de renommer cette famille UGM, pour Universal Genome Maintenance. Cependant, aucun modèle clair ne permettait d'expliquer la pléiotropie des phénotypes observés et la diversité des mécanismes dans lequel elles semblent impliquées. La question de la fonction et de l'activité de cette famille de protéines restait donc encore en suspens.

#### 4 - Kae1, l'autre enzyme de biosynthèse de t<sup>6</sup>A !

Répartition universelle, mutants à phénotypes pléiotropiques, et fonction indéterminée : ces caractéristiques rappellent celles qui menèrent à l'identification de Sua5/YrdC comme étant une enzyme de la voie de biosynthèse de t<sup>6</sup>A. La présence d'un domaine structural Kae1 fusionné à un domaine YrdC chez les carbamoyl-transférases HypF, NodU, TobZ et CmcH fut le dernier indice qui mena à cette conclusion: YgjD/Kae1 apparaît comme le candidat idéal pour être la seconde enzyme manquante de cette voie métabolique, aux côtés de YrdC/Sua5.

Ce raisonnement fut mentionné lors d'un Workshop sur les ARNt en 2010 (EL YACOUBI ET AL., 2011), et sa démonstration par une approche génétique fut publiée presque simultanément l'année suivante par plusieurs équipes (SNIRIVASAN ET AL., 2011, EL YACOUBI ET AL., 2011, DAUGERON ET AL., 2011). Chez *S. cerevisiae*, il fut ainsi montré que la mutation du



gène codant la protéine Kae1 conduit une diminution, voire à l'absence de t<sup>6</sup>A au niveau des ARNt : Kae1 semble donc bien impliquée, au côté de Sua5, dans la biosynthèse de ce nucléoside modifié. Cette découverte permet enfin, 40 ans après sa découverte, l'étude *in vivo* du rôle de t<sup>6</sup>A chez les Eucaryotes : il apparaît que l'absence de cette modification, chez des mutants Kae1 et Sua5, conduit à des erreurs d'initiation au niveau des codons AUG et provoque des décalages de cadre de lecture (EL YACOUBI ET AL., 2011, DAUGERON ET AL., 2011). Ces problèmes traductionnels pourraient conduire à une production anormale de certaines protéines clés du métabolisme (riches en codons ANN), et permettraient donc d'expliquer la pléiotropie des phénotypes observés chez les mutants Sua5/YrdC et Kae1/YgjD.

## 5 - Universalité du rôle de la famille Kae1/YgjD dans la biosynthèse de t<sup>6</sup>A

Alors que pour la famille Sua5, la question de l'universalité de la fonction, à l'aide d'expériences de complémentation, semble avoir été clairement résolue (EL YACOUBI ET AL., 2009), cette même question pour la famille Kae1/YgjD semble beaucoup plus complexe et aboutit à des résultats parfois contradictoires.

La version humaine de Kae1 peut complémenter partiellement le phénotype de retard de croissance chez les mutants Kae1 de Levure, suggérant une conservation de la fonction de Kae1 chez ces deux Eucaryotes (KISSELEVA ET AL., 2006). La déplétion de YgjD chez *E. coli* et *Staphylococcus aureus* provoque une réduction significative de la quantité de t<sup>6</sup>A (SRINIVASAN ET AL., 2011) (LEI ET AL., 2012), indiquant que les orthologues bactériens sont également impliqués dans cette voie de biosynthèse.

Enfin, un mutant  $\Delta$ Kae1 chez *S. cerevisiae* présente encore une quantité faible mais détectable de t<sup>6</sup>A, alors que un double mutant  $\Delta$ Kae1 et  $\Delta$ Qri7 ne présente plus aucune trace de t<sup>6</sup>A (SNIRIVASAN ET AL., 2011) et la délétion du signal d'adressage mitochondrial de Qri7 chez le mutant  $\Delta$ Kae1 rétablit la présence de t<sup>6</sup>A au niveau des ARNt (EL YACOUBI ET AL., 2011). Ainsi, l'orthologue mitochondrial semble lui aussi intervenir dans la synthèse de t<sup>6</sup>A.



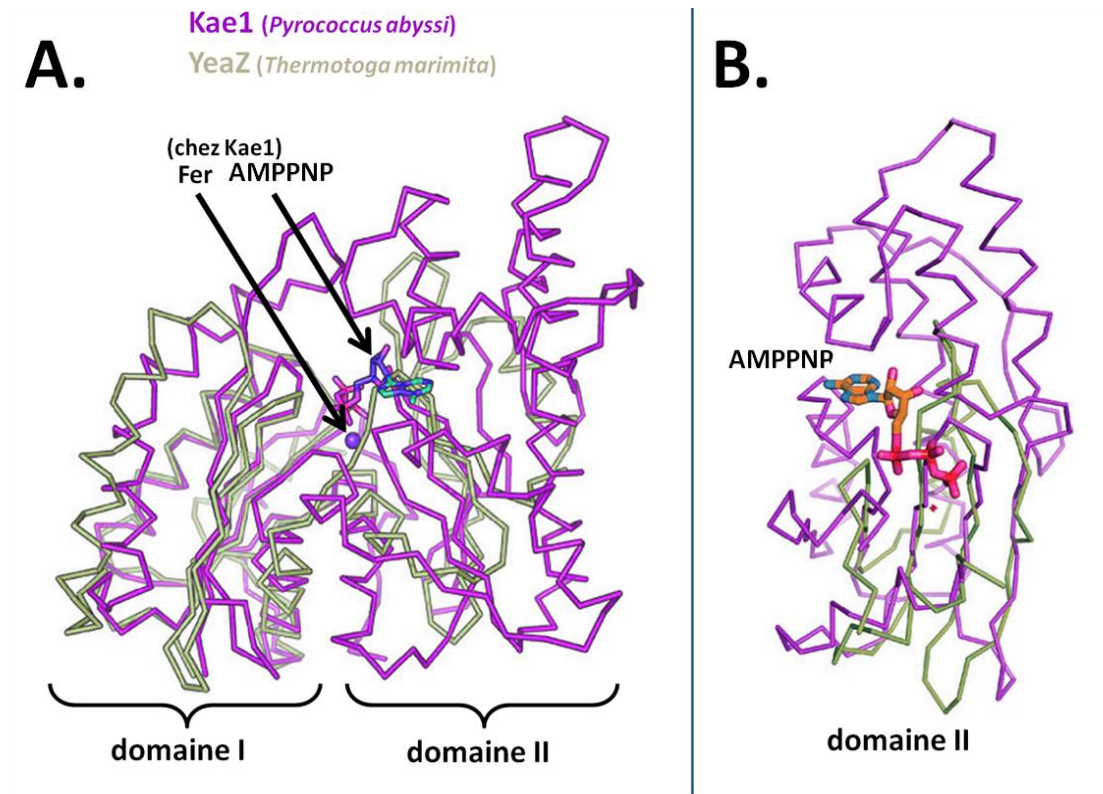


Alors que ces données suggèrent que la fonction des membres de cette famille d'orthologues est universellement conservée, d'autres expériences de trans-complémentation aboutissent à des résultats intrigants.

Chez des mutants conditionnels de YgjD de *E. coli*, l'expression de Kae1 de *S.cerevisiae* ou de l'Archée mésophile *M.maripaludis* ne permet pas de restaurer un phénotype normal (OBERTO 2009, EL YACOUBI ET AL., 2011). De manière surprenante, l'apport de YgjD de *B. subtilis*, orthologue proche de YgjD *E. coli*, n'est également pas capable de compléter la mutation. En revanche, l'apport simultané de YgjD et de son paralogue YeaZ de *B. subtilis* restaure le phénotype normal ! Cela suggère que la fonction des protéines Kae1/YgjD est dépendante d'un environnement protéique très précis et de la présence de certains partenaires, ce qui peut expliquer pourquoi il est difficile d'établir pour cette famille des trans-complémentations.

Concernant l'orthologue mitochondrial, les travaux sont contradictoires. D'une part, Oberto et al. ont montré que Qri7 complète un mutant conditionnel YgjD chez *E. coli* (OBERTO ET AL., 2009). Au contraire, El Yacoubi et al. n'obtiennent pas de complémentation de YgjD par Qri7 (EL YACOUBI ET AL., 2011). Nous discuterons de ces travaux dans ce manuscrit.

**Au bilan, ces différentes études *in vivo* montrent l'implication de la famille YgjD/Kae1 dans la voie de biosynthèse de t<sup>6</sup>A. Bien que hautement conservées, les versions Archée/Eucaryotes, Bactériennes et mitochondriales présentent certaines différences structurales, et leur activité biochimique, qui reste à établir précisément, pourrait être différente. De plus, il semble qu'un environnement cellulaire/protéique précis et spécifique soit nécessaire à leur fonction. Ainsi, de nombreuses zones d'ombre entourent encore cette famille de protéines.**



**FIGURE 40 : Mise en évidence des différences structurales entre YeaZ et Kae1.**

A. Superposition des structures de Kae1 (de *Pyrococcus abyssi*) , en violet, et YeaZ (de *Thermotoga maritima*), en gris. Les deux protéines ont un repliement caractéristique de la famille HALF, composé de deux domaines notés I et II.

Le domaine I (N-terminal) de YeaZ et Kae1 présentent la plus forte similitude. Le Fer et l'AMPPNP présent uniquement au sein de la protéine Kae1 sont mis en évidence.

B. Superposition du domaine II uniquement. YeaZ ne possède pas les éléments structuraux permettant de constituer un site de fixation au nucléotide.

d'après Xu et al, 2010

## 6 - Un puzzle complet ?

Les données issues des études de génétique concernant les familles universelles YgjD/Kae1 et YrdC/Sua5 permettent donc de poser la question suivante : ces protéines jouent-elles un rôle direct dans la biosynthèse de t<sup>6</sup>A au niveau des ARNt ?

L'étude de El Yacoubi *et al.* 2011 mentionne qu'une réaction *in vitro* de biosynthèse de ce nucléoside fut tentée sans succès avec les protéines YrdC et YgjD d'*Escherichia coli*. L'intervention d'autres acteurs protéiques, suggérée par les données décrites précédemment, fut donc proposée. Les regards se tournèrent alors naturellement vers les partenaires de YgjD/Kae1 dans les trois domaines du vivant. Il s'avère qu'effectivement, la famille Kae1/YgjD constitue un groupe de protéines très... sociables !

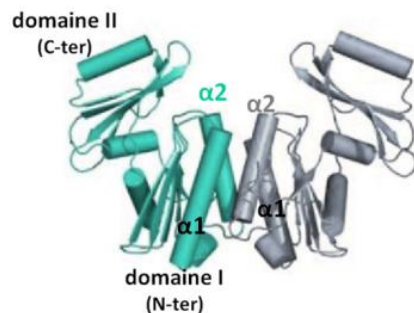
## C – Du côté des Bactéries : les partenaires de YgjD

### 1 - YeaZ : un air de famille

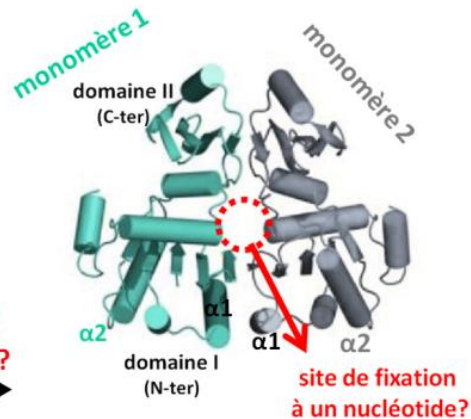
Le premier partenaire de YgjD fut identifié lors d'une étude à grande échelle des interactions entre les protéines conservées essentielles de *E. coli* (BUTLAND ET AL. 2005), confirmé par la suite lors de plusieurs études *in vivo* et *in vitro* (HANDFORD ET AL., 2009, NICHOLS ET AL., 2013): il s'agit de la protéine **YeaZ**. Ce paralogue d'YgjD (cf **FIGURE 38**), qui présente 29% d'identité avec ce dernier, est présent chez la plupart des Bactéries, à l'exception de quelques espèces de Mollicutes. Tout comme *ygjD*, *yeaZ* apparaît être essentiel chez *E. coli* (ARIGONI ET AL., 1998).

La structure de YeaZ a été résolue chez plusieurs espèces de Bactéries : *Salmonella typhimurim* (NICHOLS ET AL. 2005), *Escherichia coli* (JEUDY ET AL. 2005), *Thermotoga maritima* (XU ET AL. 2010), *Vibrio parahaemolyticus* (AYDIN ET AL., 2011). A l'image de Kae1 et de YgjD, cette protéine présente un repliement caractéristique de la famille ASHKA, et comporte deux domaines. Le domaine amino-terminal (domaine I) présente le plus de similarité au niveau séquence et structure avec Kae1/YgjD, alors que le domaine carboxy-terminal (domaine II) en diffère fortement, notamment au niveau de son orientation (**FIGURE 40**). YeaZ ne possède pas les sites de liaison à l'ATP et au métal, suggérant que cette protéine possède une activité différente de son paralogue YgjD.

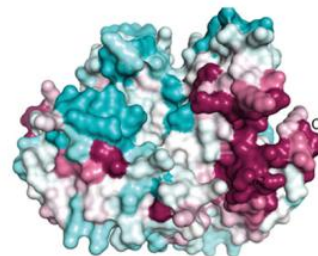
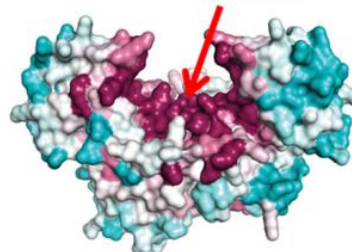
**A.**      **Forme dimérique 1**  
              **(A<sub>2</sub>)**



### Forme dimérique 2 (AB)



**B.**



peu conservé  très conservé

**FIGURE 41:** Comparaison des formes homodimériques de la protéine YeaZ de *Thermotoga maritima*

A. Mise en évidence de la structure des homodimères. Un des deux monomères est coloré en bleu, l'autre est coloré en gris. Les domaines I et II sont indiqués pour un monomère. Les hélices  $\alpha 1$  et  $\alpha 2$ , impliqués dans l'interaction de la forme dimérique 1, sont indiquées.

B. Mise en évidence des régions conservées, d'après la comparaison de 237 séquences d'orthologues de YeaZ.

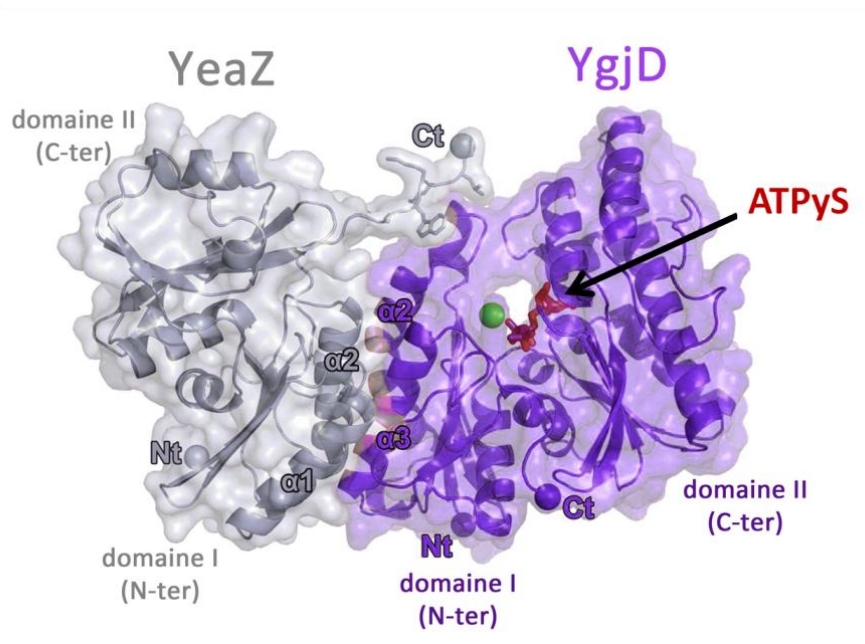
Les structures de YeaZ de *Vibrio parahaemolyticus*, *Salmonella typhimurium*, *Escherichia coli* ont toutes été résolues sous la forme dimérique 1, appelé également A2. Le mode d'interaction de ce dimère correspond au mode d'interaction observé dans la structure de YeaZ/YgjD (Nichols *et al.* 2013), et est indiqué en rouge. La forme dimérique 2, appelée également AB, n'a été observée que chez YeaZ de *Thermotoga maritima*. Le site de fixation à un nucléotide potentiel créé au niveau de l'interface est indiqué en rouge.

La dynamique de ces deux formes dimériques pourrait dépendre de la présence de nucléotides ou de partenaires et correspondre à un mécanisme de « switch » moléculaire (d'après Nichols *et al.*, 2006, Xu *et al.* 2010, Aydin *et al.* 2011, Nichols *et al.*, 2013)

*In vitro*, YeaZ se présente sous forme d'homodimère, tout comme YgjD, et toutes les structures cristallines citées précédemment ont été résolues sous une configuration dimérique appelée forme 1/A2 (FIGURE 41). Dans cette configuration, l'interface entre les deux monomères correspond aux régions les plus conservées de la protéine et fait intervenir les hélices  $\alpha_1$  et  $\alpha_2$  du domaine I. Ces données structurales montrent également des différences interspécifiques au niveau de cette interface de dimérisation: elle diffère légèrement selon l'espèce d'origine. De plus, il a été montré que l'interaction entre YgjD et YeaZ d'une même espèce est plus forte qu'une interaction entre YgjD et YeaZ issues de deux espèces différentes (RAJAGOPALA ET AL, 2011). Enfin, la délétion du gène *ygjD* chez *E. coli* ne peut être complétée que par l'apport simultané du couple YeaZ/YgjD de *Bacillus subtilis* (EL YACOUBI ET AL, 2011). D'un point de vue évolutif, ces données suggèrent que le couple YgjD/YeaZ ait évolué à partir d'un homodimère, l'une des deux sous-unités (YeaZ), aurait divergé pour acquérir une nouvelle fonctionnalité ; et le couple YeaZ/YgjD aurait alors co-évolué dans chaque espèce, formant des interfaces spécifiques. La fonction et l'activité de ces deux protéines apparaît donc être intimement liée.

Récemment, la structure de l'hétérodimère YgjD/YeaZ de *Salmonella typhimurium* a été décrite (NICHOLS ET AL, 2013), montrant une configuration et une interface similaire à cette forme 1/A2 (FIGURE 42). De manière intéressante, une seconde configuration homodimérique de YeaZ, appelée forme 2/AB a été décrite chez *Thermotoga maritima* (XU ET AL, 2010), où les deux monomères interagissent selon une interface différente et pourraient former un nouveau site de fixation à un nucléotide (FIGURE 41). De plus, certains résidus conservés hydrophobes se trouvent exposés au solvant dans cette configuration, pouvant constituer une nouvelle plate-forme d'interaction. Il a donc été suggéré que ce second état conformationnel pourrait traduire un mécanisme de « switch » moléculaire (c'est-à-dire un basculement d'une conformation à l'autre), modulé par la fixation ou l'hydrolyse de l'ATP, et qui contrôlerait l'association entre YeaZ, YgjD et/ou d'éventuels partenaires (AIDYN ET AL, 2011).

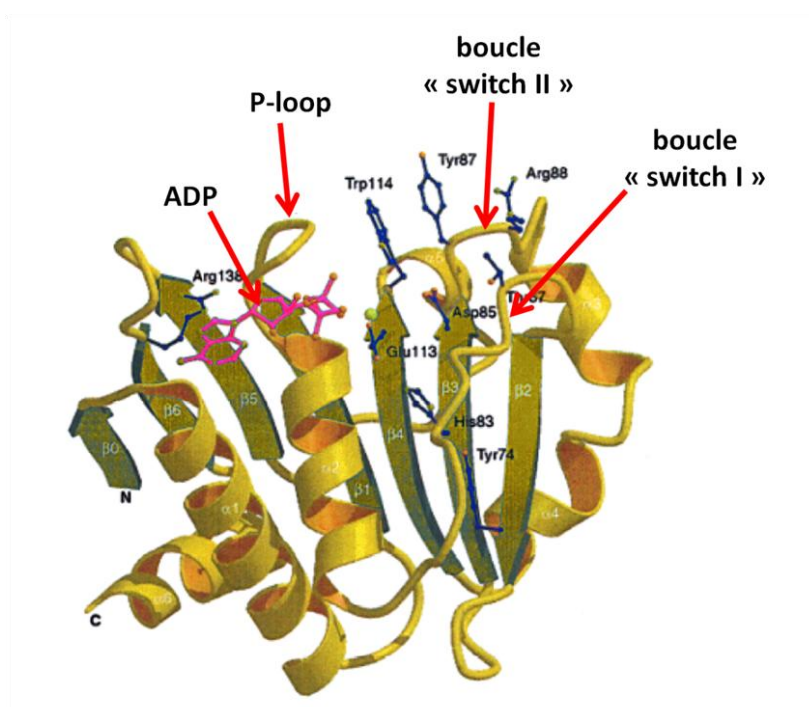




**FIGURE 42:** Structure de l'hétérodimère YeaZ-YgjD de *Salmonella typhimurium*, cristallisé en présence d'ATPyS

La protéine YeaZ est colorée en gris, la protéine YgjD en violet.

La molécule d'ATPyS en rouge, et l'atome de Magnésium présent au sein du site actif en vert. L'atome de Zinc, présent également au sein du site actif de YgjD, n'est pas représenté. Les hélices  $\alpha$  impliquées dans l'interface entre les deux protéines et les extrémités N et C terminales sont indiquées. (d'après Nichols *et al.* 2013)



**FIGURE 43:** Structure de la protéine YjeE d'*Haemophilus influenzae*.

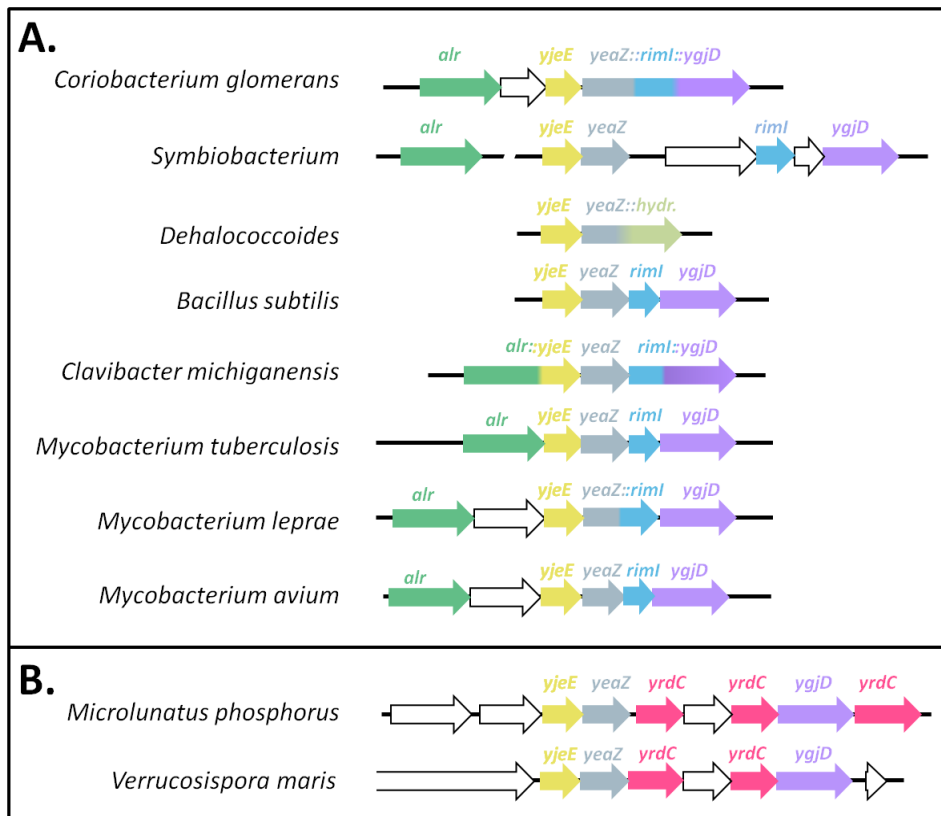
Le nucléotide ADP est coloré en rose. Les acides aminés conservés sont mis en évidence (en bleu), et les régions conservées P-loop et boucles « switch » I et II sont mises en évidence. (d'après Teplyakov *et al.*, 2002)

## 2 - YjeE : un interrupteur moléculaire ?

Une troisième protéine complète ce couple : la protéine YjeE. Absent chez les Archées et les Eucaryotes, tout comme YeaZ, le gène *yjeE* est hautement conservé chez les Bactéries mais absent notamment chez certaines espèces de Mollicutes (TEPLYAKOV ET AL 2002). Le gène *yjeE* et ses orthologues représentent un exemple de gène dont l'essentialité est discutée. En raison de l'impossibilité d'obtenir une délétion de ce gène chez *B. subtilis* et *E. coli*, il a d'abord été défini comme essentiel (ALLALI HASSANI ET AL., 2004, FREIBERG ET AL., 2001). Cependant, des études plus récentes montrent la possibilité d'obtenir un mutant délété viable chez *B. subtilis* (HUNT ET AL., 2006), alors qu'une dernière étude chez *E. coli* le requalifia d'essentiel (HANDFORD ET AL., 2009). Au bilan, malgré ces contradictions, il semble que YjeE joue un rôle fondamental au niveau du fonctionnement cellulaire.

YjeE est une protéine de 17kDa chez *E. coli* qui présente deux motifs Walker A (ou P-loop pour Phosphate binding loop) et B conservés, ce qui la classe parmi la super famille des ATPase à boucle-P (SARASTE ET AL, 1990 ; TEPLYAKOV ET AL, 2002). *In vitro*, elle a été observée sous forme de monomère, de dimère et d'oligomère. L'état dimérique a été observé chez différentes espèces (*E. coli*, *B. subtilis*), il a été suggéré que cet état puisse jouer un rôle dans la régulation de la fonction de cette protéine (KARST ET AL, 2009). La résolution de la structure de YjeE d'*Haemophilus influenzae* (TEPLYAKOV ET AL, 2002) a fourni plusieurs informations particulièrement intéressantes quant au fonctionnement possible de cette protéine (FIGURE 43). D'une part, elle révèle une topologie caractéristique des ATPases à P-loop, mais qui présente un repliement inédit parmi cette grande famille de protéines, faisant d'YjeE le représentant d'une nouvelle sous-famille d'ATPase. Une molécule d'ADP est co-cristallisée au sein du site actif, et seule la présence d'ADP ou d'ATP+Mg<sup>2+</sup> permet la cristallogénèse.

Cela indique que YjeE doit présenter des conformations qui diffèrent en fonction de la présence ou l'absence de nucléotide. Cette interprétation est confortée par la présence, en plus de la boucle P, de deux boucles hautement conservées (appelées «switch»), que l'on retrouve dans les protéines à P-loop qualifiées d'« interrupteurs moléculaires». Chez ces protéines, l'hydrolyse du nucléotide provoque au niveau du site actif un changement conformationnel qui est amplifié et transféré à des parties distales de la molécule.



**FIGURE 44 :** Contexte génomique des gènes *yjeE*, *yeaZ*, *ygdD* et *yrdC* chez certaines espèces de Bactéries

**A.** Association entre les gènes *yjeE*, *yeaZ*, *rimI* et *ygdD*

**B.** Espèces où l'on retrouve le gène *yrdC* associé aux gènes *ygdD*, *yjeE*, *yeaZ* et *ygdD* d'après Oberto et al., 2013.

Les synténies ont été réalisées avec le logiciel SYNTAX (Oberto et al., 2013). Les gènes fusionnés sont indiqués par un :

*rimI*= N acétyl transférase de la protéine ribosomale S18, *alr*= alanine racémase, impliquée dans la synthèse du peptidoglycane, *hydr*= hydrolase

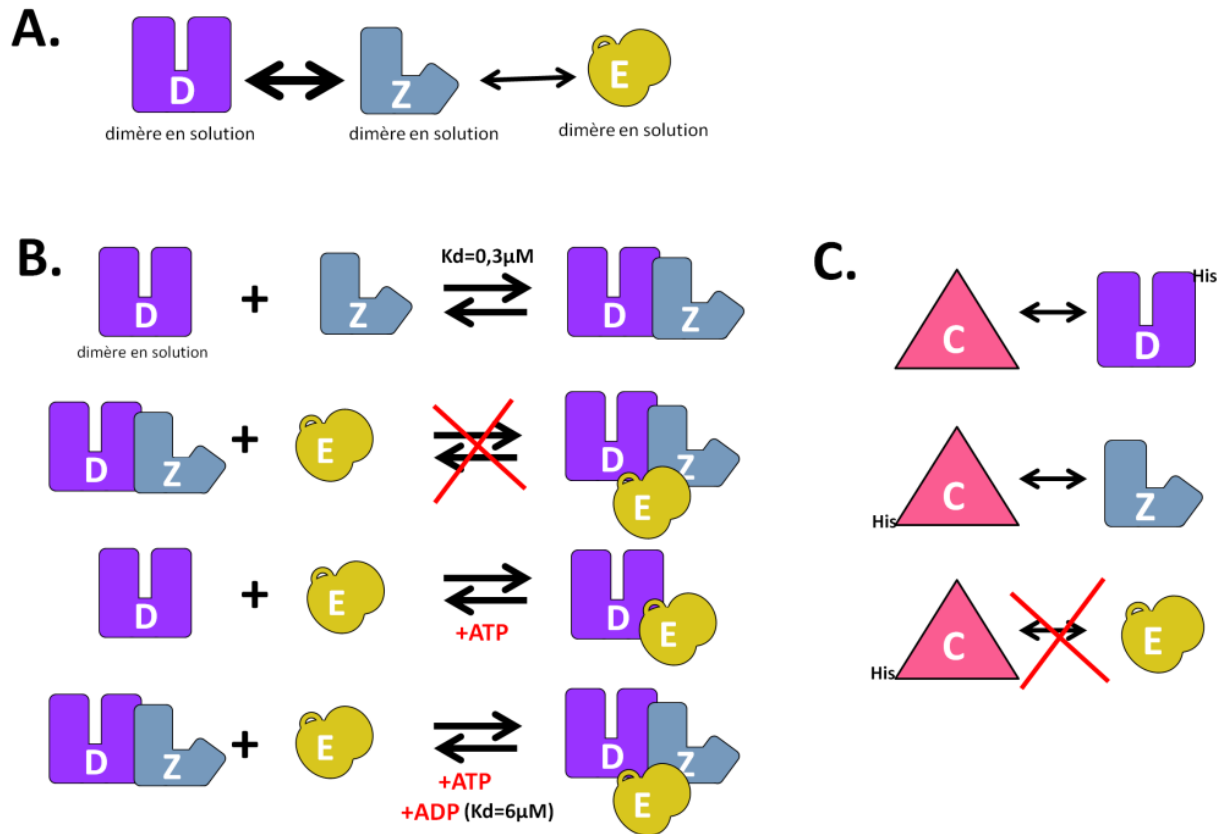


Dans l'une des boucles « *switch* » d'YjeE, on retrouve une paire de résidus aromatiques (Tyrosine 87 et Tryptophane 114), accompagnée d'un résidu basique (Arginine 88) qui sont hautement conservés. Ces résidus, que l'on retrouve également chez certains facteurs de transcription, amino-acyl ARNt transférases et ribonucleoprotéines (TEPLYAKOV ET AL, 2002), pourraient constituer un domaine de liaison à un acide nucléique une fois exposés au solvant.

L'ensemble de ces données suggèrent donc que YjeE fonctionne comme un interrupteur moléculaire via son activité ATPase. Cependant cette activité *in vitro* est très faible, quel que soit l'organisme d'origine, avec un  $k_{cat}$  de l'ordre de  $1h^{-1}$  à  $10h^{-1}$  (autrement dit, une molécule d'YjeE hydrolyse une molécule d'ATP en 1 à 10h !) (TEPLYAKOV ET AL, 2002, HASSANI ET AL 2004, KARST ET AL, 2009, HANDFORD ET AL, 2009), et la protéine présente une meilleure affinité pour l'ADP (produit d'hydrolyse) que pour l'ATP. Pour ces raisons, il a été suggéré que l'activité ATPase de YjeE et le phénomène de « *switch* » moléculaire qui pourrait y être associé dépendent probablement de l'interaction avec un substrat ou un partenaire spécifique non identifié lors de ces études.

### 3 - Ménage à trois...

De manière intéressante, chez de nombreuses espèces de Bactéries, les gènes *yjeE*, *yeaZ* et *ygjD* sont physiquement adjacents sur le chromosome, selon un ordre (ou synténie) conservé, suggérant un possible lien physique et fonctionnel (FIGURE 44) (HANDFORD ET AL, 2009, OBERTO ET AL, 2013). Ces liens furent confirmés par des expériences d'interaction en double-hybride réalisées chez *E. coli* (HANDFORD ET AL, 2009), montrant que YeaZ interagit d'une part avec YgjD, et d'autre part avec YjeE. En revanche, aucune interaction n'a été détectée entre YjeE et YgjD par cette méthode. De plus, la sur-expression de YgjD chez *E. coli* semble abolir l'interaction YeaZ/YjeE observée. Ainsi, YeaZ pourrait interagir alternativement avec YgjD d'une part et YjeE d'autre part. (FIGURE 45-A). Récemment, la dynamique de ce trio a été étudiée *in vitro* par microcalorimétrie (ITC, pour *isothermal titration calorimetry*) (NICHOLS ET AL., 2013). Alors que les protéines YgjD et YeaZ forment spontanément un hétérodimère, la protéine YjeE n'interagit avec YgjD ou le dimère préformé YgjD/YeaZ qu'en présence de nucléotide, ATP ou ADP. Dans ce dernier cas, la constante de dissociation mesurée pour le complexe ternaire YgjD/YeaZ/YjeE est de  $6\mu M$  (FIGURE 45-B).



**FIGURE 45: Modèles d'interaction entre les protéines YgjD, YeaZ, YjeE et YrdC**

A. D'après Handford *et al*, 2009, par expérience de double hybride et de co-purification. YeaZ (notée Z) interagit avec YgjD (notée D) et avec YjeE (E), de manière exclusive. L'interaction entre ZD apparaît plus stable que l'interaction YjeE (représentée par des flèches d'épaisseur différente)

B. D'après Nichols *et al*, 2013, par expérience de microcalorimétrie. D et Z forment spontanément un hétérodimère, avec une constante de dissociation de 300nM. E n'interagit pas avec ce dimère en l'absence de nucléotide, mais peut interagir avec D ou DZ en présence d'ATP ou d'ADP, pour former un complexe multiprotéique. La constante d'association a pu être mesurée dans le cas de la présence d'ADP et est égale à 6μM.

C. D'après Deutsch *et al*, 2012, par expérience de co-purification sur colonne. La présence de l'étiquette hexa-histidine utilisé lors de ces expériences est indiquée par le symbole His. La protéine D-His retient sur colonne de nickel la protéine YrdC (notée C), et la protéine C-His retient sur colonne la protéine Z. En revanche, la protéine C-His ne retient pas la protéine E.

Un tel comportement pourrait traduire le phénomène de *switch* moléculaire prévu pour YjeE. Cependant, un complexe à trois protéines n'a pour le moment pas été isolé physiquement. De plus, aucune interaction n'a pu être établie entre YjeE et YeaZ en l'absence de YgjD, en contradiction apparente avec les données de Handford *et al.*

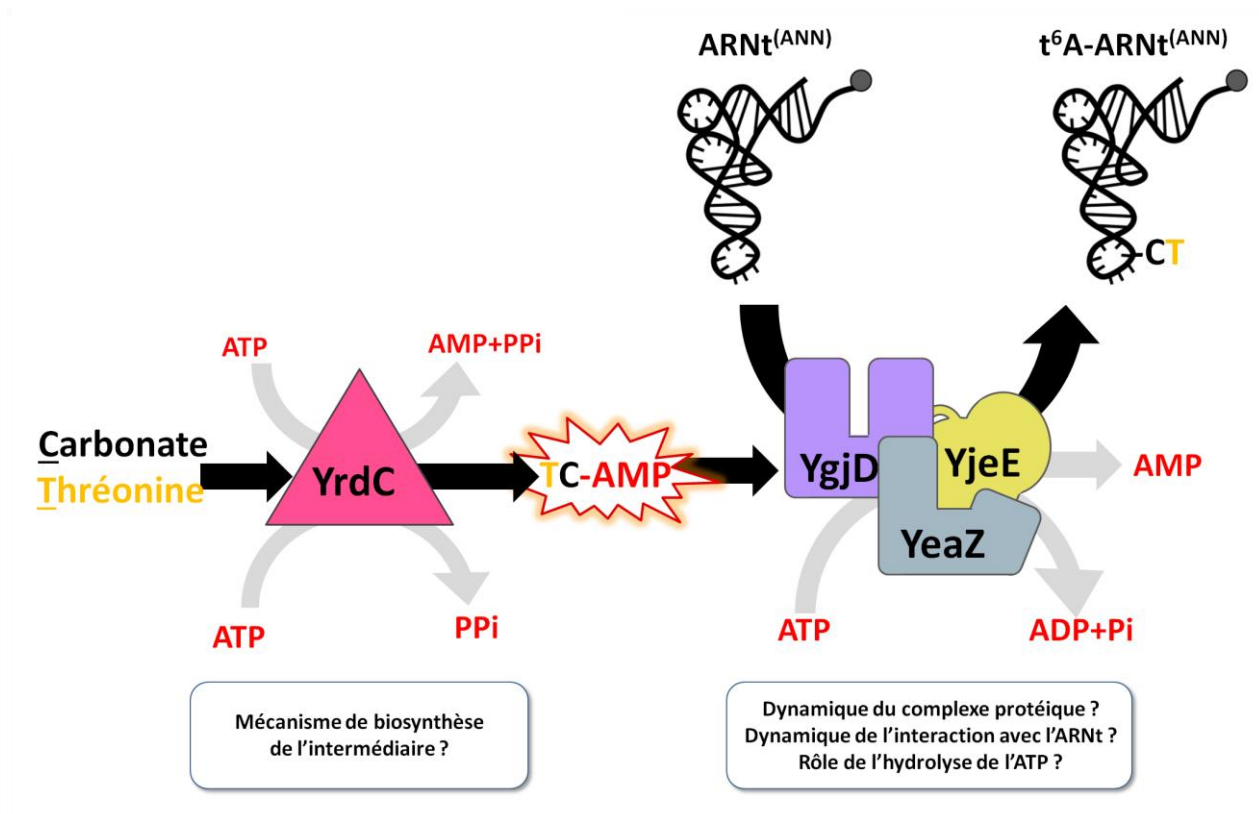
Il semble donc que YgjD, YeaZ et YjeE forme un réseau de protéines dynamique encore largement incompris, contrôlé par la présence de nucléotide, et faisant peut-être intervenir des phénomènes de « *switch* » moléculaire entre les différents partenaires. Des études plus complètes sont ainsi nécessaires pour comprendre de manière précise les modalités de ces interactions.

#### 4 - La fin du mystère... chez les Bactéries !

Après avoir démontré par génétique que les protéines Sua5/YrdC et Kae1/YgjD étaient impliquées dans la voie de biosynthèse de  $t^6A$ , il restait à démontrer leur implication directe par des méthodes biochimiques. N'ayant pas pu reconstituer cette réaction de synthèse *in vitro* avec les protéines YrdC et YgjD purifiées de *E. coli* (EL YACOUBI ET AL, 2011), il fut donc naturellement suggéré, au vu des observations précédentes, que les partenaires YeaZ et YjeE pouvaient faire partie de cette voie enzymatique.

Ainsi, en 2012, le résultat tant convoité fut enfin publié : la biosynthèse de  $t^6A$  *in vitro* à partir de protéines purifiées de *E. coli* (DEUTSCH ET AL, 2012). Comme il était attendu, cette biosynthèse requiert la présence simultanée des 4 protéines : YrdC, et le trio YgjD, YjeE et YeaZ. De manière intéressante, une interaction directe entre YrdC et respectivement YgjD et YeaZ a été mise en évidence lors de cette étude par expérience de rétention sur colonne (*pull down* en anglais) (FIGURE 45-C). En raison de cette implication directe dans cette voie de biosynthèse, les auteurs ont proposé de renommer les protéines YrdC, YgjD, YjeE et YeaZ en TsaC (pour *t*-*s*ix-*A*), TsaD, TsaE et TsaB (le B remplaçant le Z de YeaZ).

Ensemble, ces quatre protéines incorporent carbonate et thréonine au sein d'ARNt transcrits *in vitro*, sous la forme de thréonylcarbamoyl-adénosine ( $t^6A$ ), en présence d'ATP et de différents ions. De plus, seuls les ARN<sup>t<sup>Lys</sup></sup> et ARN<sup>t<sup>Thr</sup></sup> (porteur d'anticodon de la forme NNU)



**FIGURE 46 :** Modèle de biosynthèse de t<sup>6</sup>A chez les Bactéries par les protéines YgjD, YeaZ, YjeE et YrdC.

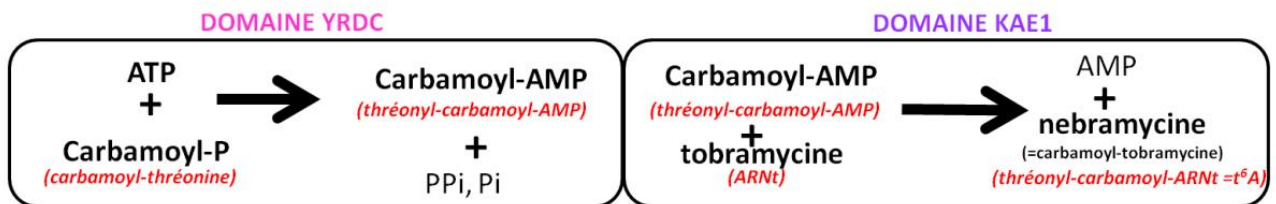
Les différentes activités associées aux protéines sont représentées. TC-AMP = thréonylcarbamoyl-AMP. ANN= séquence du codon lu par l'ARNt substrat. N= U, G, C ou A. d'après Deutsch *et al.* 2012, Lauhon *et al.*, 2012.

sont modifiés, ce qui est en accord avec les données sur la distribution naturelle de t<sup>6</sup>A. Le fait de pouvoir utiliser comme substrats des transcrits *in vitro* indique que la présence de modifications préalables sur l'ARNt n'est pas nécessaire à la reconnaissance du substrat par ces protéines. En revanche, un ARN synthétique de 17 nucléotides reproduisant la boucle anticodon de l'ARNt n'est pas modifié, suggérant que cette machinerie bactérienne de synthèse de t<sup>6</sup>A requiert probablement des déterminants présents dans la structure entière de l'ARNt, à l'image de la machinerie Eucaryote (d'après les travaux sur les ovocytes de Xénope, MORIN ET AL., 1998).

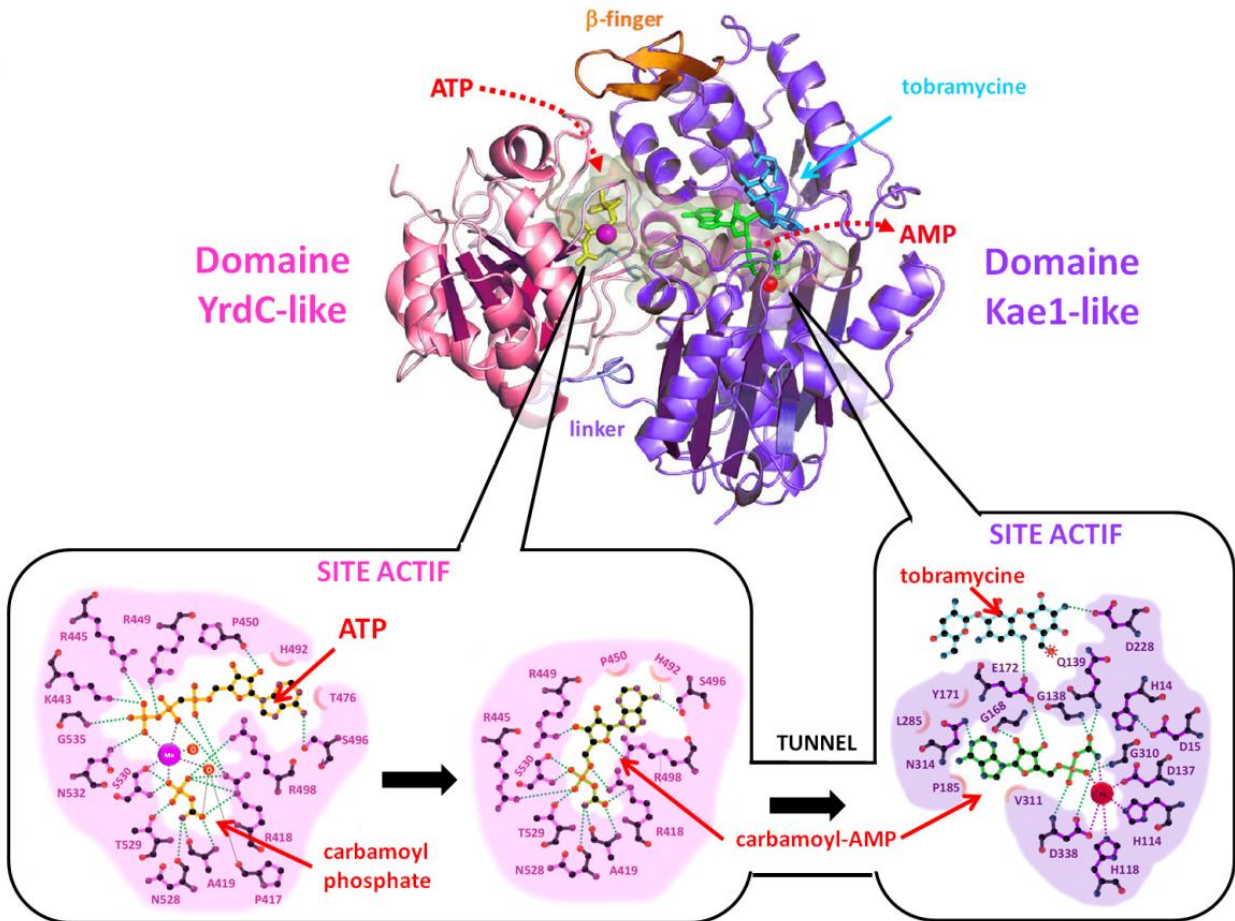
Dans les conditions de cette réaction, la protéine YrdC hydrolyse l'ATP en AMP en présence de thréonine et de carbonate. D'autre part, la présence simultanée des protéines YgjD, YeaZ et YjeE provoque l'hydrolyse de l'ATP en ADP, indépendamment de la présence d'autres facteurs (carbonate, thréonine ou ARNt). Ces données ont été complétées récemment par une étude de ce système enzymatique *in vitro* à partir des protéines de *Bacillus subtilis* (LAUHON ET AL. 2012), révélant certains détails mécanistiques. La protéine YrdC produit, à partir d'ATP, de thréonine et de carbonate, un intermédiaire réactionnel sous forme de carbamoyl-thréonine-adénylate (TC-AMP). Malgré une durée de vie limitée (entre 3,5mn et 16mn à 37°C) cet intermédiaire a pu être isolé par HPLC. Il a alors été montré que le trio YgjD/YeaZ/YjeE transfère le groupement thréonylcarbamoyl de cet intermédiaire sur l'ARNt pour former t<sup>6</sup>A, et la présence et l'hydrolyse de l'ATP ne semble pas nécessaire lors de cette étape.

Au bilan, la biosynthèse de t<sup>6</sup>A chez les Bactéries se déroule en deux étapes principales, représentée dans la **FIGURE 46**. Cependant, de nombreuses inconnues restent encore à élucider au niveau de ce mécanisme.

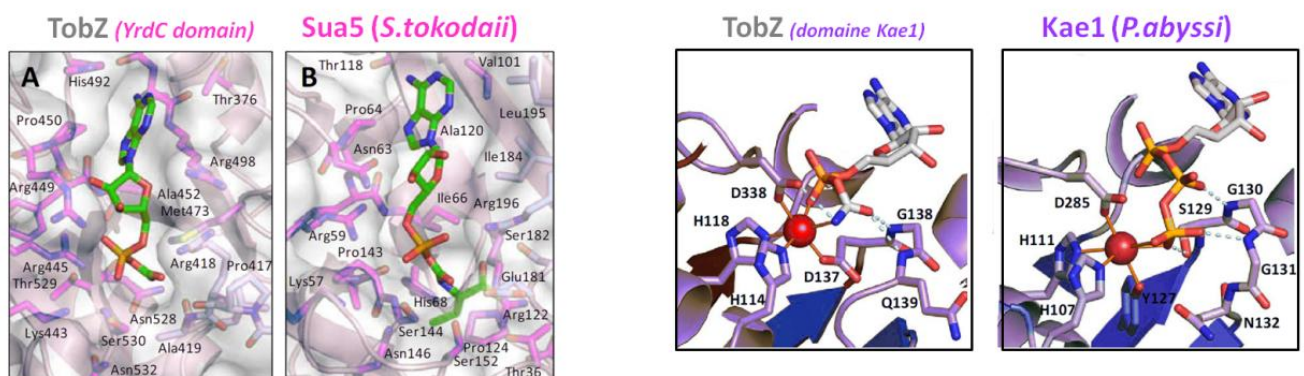
A.



B.



C.



**FIGURE 47 : Structure et mécanisme catalytique de la protéine TobZ, illustrant le rôle des domaines YrdC et Kae1.**

**A. Mécanisme catalytique de la réaction de carbamoylation par la protéine TobZ.**

Les deux étapes catalysées respectivement par les domaines YrdC (en rose) et Kae1 (en violet) sont représentées en deux cadrans séparés. En rouge entre parenthèse sont indiqués les équivalents proposés pour la réaction de biosynthèse de t<sup>6</sup>A.

[suite de la légende sur page de figure suivant]



**Comment est synthétisé l'intermédiaire TC-AMP par YrdC ?** Les mesures quantitatives de production d'AMP semblent indiquer qu'une seule molécule d'ATP est requise, et suggère que la condensation du carbonate avec la thréonine soit ATP-indépendante, mais ces données restent à confirmer (LAUHON ET AL., 2012)

**D'autre part, comment l'intermédiaire est-il transféré après sa production?** Le découplage de la réaction par isolation de l'intermédiaire en HPLC montre que ce dernier peut être libéré dans le solvant, cependant cela n'exclut pas la possibilité qu'il soit transféré directement de YrdC aux autres protéines par interaction directe.

**Et quel est le rôle de chacune des trois protéines YgjD, YjeE et YeaZ?** Chacune des protéines possède un site potentiel de fixation à un nucléotide : chacune d'entre elle est donc candidate à la fois pour l'hydrolyse de l'ATP observée *in vitro* mais aussi la fixation de l'intermédiaire et son transfert sur l'ARNt.

**Quelle est l'importance et le rôle de l'hydrolyse de l'ATP par ces protéines ?** D'après les connaissances structurales concernant YjeE et YeaZ, on peut s'attendre à ce que l'hydrolyse de l'ATP qui semble accessoire à la seconde étape de la réaction, soit liée à des changements conformationnels jouant un rôle dans la formation d'un complexe ternaire, dans l'efficacité de la catalyse et/ou dans la reconnaissance du substrat ARNt.

## 5 - Les protéines TobZ et HypF: une mine d'information!

Quelques unes de ces questions trouvent en partie réponse dans l'étude des fameuses carbamoyl-transférases présentées dans un chapitre précédent, qui servirent de piste initiale dans la recherche de la t<sup>6</sup>A synthétase. L'une d'entre elle, TobZ, est une O-carbamoyl-transférase qui a pour substrat la tobramycine, un aminoglycoside antibiotique, et qui possède un domaine structural YrdC en position N-terminal fusionné à un domaine structurale Kae1 en position C-terminale (comme les protéines NodU et CmCH) (cf **FIGURE 35**). Récemment, la résolution de la structure de la protéine TobZ de la bactérie *Streptoalloteichus tenebrarius* a révélé le mécanisme d'action de cette enzyme et le rôle respectif des domaines structuraux YrdC et Kae1 (PARTHIER ET AL., 2012) dans la réaction de carbamoylation, permettant de dresser des parallèles avec la réaction de biosynthèse de t<sup>6</sup>A (**FIGURE 47**). La résolution récente de la structure et du mécanisme catalytique de HypF, une

**FIGURE 47 – SUITE DE LA LEGENCE** Structure et mécanisme catalytique de la protéine TobZ, illustrant le rôle des domaines YrdC et Kae1

**B. Structure de la protéine TobZ de la bactérie *Streptoalloteichus tenebrarius* avec représentation schématique de la réaction.** L'atome de manganèse retrouvé dans le domaine YrdC est indiqué en rose, l'atome de fer dans le domaine Kae1 en rouge. La molécule de tobramycine (le substrat carbamoylé) en bleu, l'intermédiaire AMP-carbamoyl en jaune dans le domaine YrdC, en vert dans le domaine Kae1. Le tunnel formé par l'interface des deux domaines est coloré en gris clair.

Les encarts inférieurs visualisent les sites actifs des deux domaines. A gauche, le site actif de YrdC avant réaction, où se trouve le carbamoyl-phosphate, l'ATP (en orange) et l'atome de manganèse (en rose). Les liaisons sont représentées en segments bleu clair. A droite, après réaction, où se trouve l'intermédiaire. Le tunnel où transite l'intermédiaire entre les deux domaines est schématisé. L'encart de droite correspond au site actif de Kae1.

**C. Comparaison des sites actifs de TobZ(YrdC) avec le site actif de Sua5 de *S.tokodaii*, et de TobZ(Kae1) avec le site actif de Kae1 de *P. abyssi*.**

Les ligands respectifs trouvés dans les différents domaines sont représentés (respectivement AMP-C, AMP-TC, AMP-C et AMPPNP).

d'après Parthier *et al*, 2011



autre carbamoyltransférase fournit des informations comparables (PETKUN ET AL., 2011, SHOMURA ET AL., 2012).

Le domaine YRDC de TobZ et HypF, contenant respectivement un atome de manganèse, fixe une molécule de carbamoyl-phosphate (un carbamate dans le cas de HypF) et une molécule d'ATP, pour produire un intermédiaire Carbamoyl-AMP. Cet intermédiaire est transféré au domaine Kae1 par un tunnel formé à l'interface des deux domaines. Ainsi, dans le cas de la synthèse de  $t^6A$ , l'intermédiaire thréonylcarbamoyl-AMP produit par YrdC pourrait transiter de la même manière dans un tunnel formé par une possible interface entre YrdC et YgjD. Le site actif du domaine Kae1 de TobZ, qui contient un atome de fer (comme la protéine Kae1 de *P. abyssi*), forme alors des liaisons nécessaires pour la fixation de l'intermédiaire et fixe également le substrat tobramycine, puis transfère le groupement carbamoyl sur ce substrat, sans avoir recours à une hydrolyse d'ATP. Ainsi, de manière analogue, la protéine Kae1/YgjD serait le lieu de fixation du TC-AMP et du transfert du groupement thréonylcarbamoyl sur son substrat, l'ARNt.

La mise en évidence d'un intermédiaire adénylé au cours de ces réactions de carbamoylation a permis de révéler une information essentielle concernant la protéine Sua5 d'Archée. En effet, lors de la résolution de la structure de la protéine Sua5 de *Sulfolobus tokodaii*, une molécule d'AMP fut identifiée au sein du site actif de la protéine (AGARI ET AL., 2008). Parthier *et al.*, après avoir ré-examiné ces données, montrent que la densité électronique observée correspond en fait à une molécule de thréonyl-carbamoyl AMP, l'intermédiaire produit par YrdC de *B. subtilis*. Cette interprétation est confortée par une autre structure de Sua5, qui présente une molécule d'AMPPNP et une molécule de L-thréonine au sein de son site actif (KURATANI ET AL., 2011)

Ainsi, chez les Archées (et probablement chez les Eucaryotes), la protéine Sua5 produirait le même intermédiaire TC-AMP que son homologue bactérien; qui serait par la suite transféré au site actif de Kae1, qui à son tour transférerait le groupement carbamoyl-thréonine sur l'ARNt.





## Le mystère de la biosynthèse de t<sup>6</sup>A serait-il donc résolu, et l'affaire classée ?

Hélas (ou heureusement) non ! Car malgré ces nombreuses données, tant biochimiques que structurales, de nombreux aspects restent encore non résolus. Notamment, le rôle des protéines YjeE et YeaZ n'est toujours pas compris. Leur présence indispensable dans la machinerie de synthèse de t<sup>6</sup>A chez *E. coli* et *B. subtilis* soulève plusieurs grandes questions auxquelles je tenterai de répondre :

La mitochondrie possède des ARNt qui portent la modification t<sup>6</sup>A, mais seul un orthologue de YgjD, la protéine Qri7, a été identifié au sein de cet organite. Comment est alors réalisée la synthèse de t<sup>6</sup>A au niveau mitochondrial ?

De façon similaire, chez les Eucaryotes et les Archées, YjeE et YeaZ sont absentes : Sua5 et Kae1 peuvent-elles catalyser seules cette réaction, ou, comme chez les Bactéries, des partenaires spécifiques sont-ils requis ?

Afin de pouvoir répondre à cette dernière question, nous allons donc nous intéresser dans un dernier temps aux partenaires connus de Kae1 chez les Eucaryotes et chez les Archées.



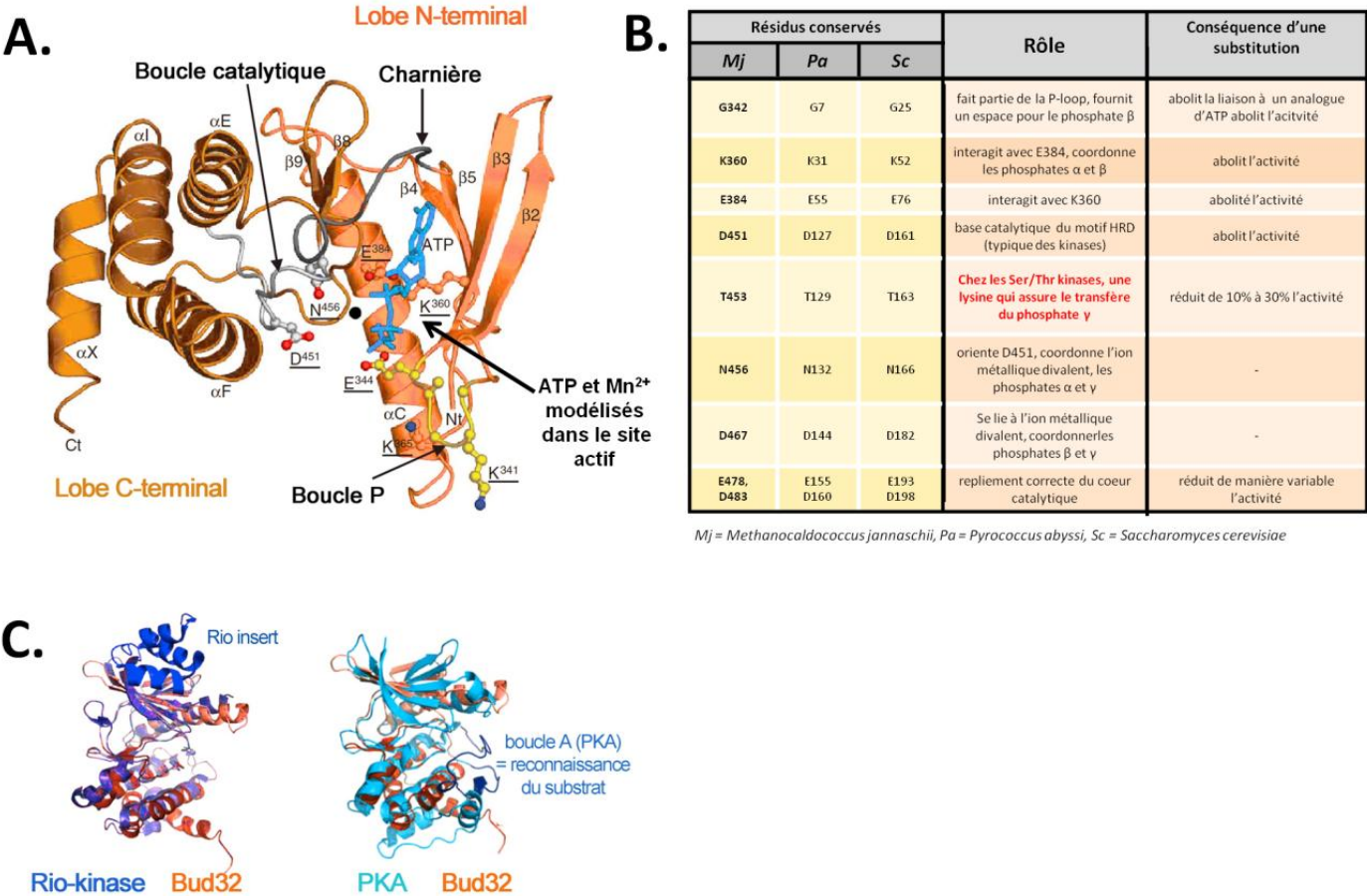
## D - Du côté des Archées et des Eucaryotes : les partenaires de Kae1

En 2004, *Lopreiato et al.* recherchent chez *Saccharomyces cerevisiae* les partenaires potentiels d'une petite sérine-thréonine kinase, appelée Bud32. Ils détectent alors par double-hybride l'orthologue de YgjD, qu'ils nomment Kae1 pour « Kinase-Associated Endopeptidase ». Avant de revenir sur l'interaction entre Kae1 et Bud32, nous allons aborder les caractéristiques de cette protéine.

### 1- L'intrigante Bud32, une kinase atypique

Identifiée chez *Saccharomyces cerevisiae* en 1997 comme étant une sérine thréonine (Ser/Thr) kinase putative (CLEMENTE ET AL., 1997), Bud32 fut rapidement l'objet d'une grande attention, pour plusieurs raisons. D'une part, sa délétion cause une altération sévère de la croissance, suggérant un rôle fondamental dans la cellule. Deuxièmement, au niveau de sa séquence, certains des motifs et domaines caractéristiques de la famille des kinases sont fortement altérés. De plus, sa taille (261 acides aminés, soit environ 30kDa chez *S. cerevisiae*) en ferait la plus petite kinase décrite (STOCCHETTO ET AL., 1997). Bud32 est une protéine hautement conservée : elle est présente virtuellement aussi bien chez tous les Eucaryotes que chez les Archées, mais absente chez les Bactéries. Il fut ainsi suggéré que Bud32 représente une famille très ancienne de kinases, dont les caractéristiques refléteraient l'état ancestral des kinases eucaryotes modernes (LEONARD ET AL., 1998, FACCHIN ET AL., 2002), et qui serait apparu avant la divergence entre Archée et Eucaryotes.

Les études biochimiques *in vitro* réalisées sur Bud32 issue de *S. cerevisiae* montrent que cette protéine présente effectivement une activité kinase : elle s'autophosphoryle et peut phosphoryler certaines protéines, comme la caséine, l'ostoponine, la glutaredoxine (STOCCHETTO ET AL 1997, LOPREIATO ET AL, 2004). Cependant, cette activité kinase *in vitro* est faible (son  $k_{cat}$  est de l'ordre de  $0,03\text{min}^{-1}$  pour les différents substrats non spécifiques utilisés, FACCHIN ET AL., 2002). Cette activité dépendante par la présence de manganèse (STOCCHETTO ET AL, 1997), ce qui n'est pas typique des Ser/Thr kinases. Ces caractéristiques biochimiques sont retrouvées chez les orthologues Bud32 des Archées *Methanocladococcus jannaschii* et *Pyrococcus abyssi* (HECKER ET AL., 2008, HECKER ET AL., 2009).



**FIGURE 48:** Propriétés structurales de Bud32 de *M.jannaschii* et comparaison avec d'autres kinases

**A.** Structure du domaine Bud32 de la protéine MJ1130. Le domaine Kae1, lié au niveau de l'extrémité N-terminale du domaine Bud32 par un linker, n'est pas représenté pour plus de clarté.

**B.** Résidus conservés du site actif et rôles déterminés par étude mutationnelle de Bud32 chez *S. cerevisiae*

**C.** Superposition des structures de Bud32 (en orange) avec Rio1 (en bleu foncé) et la Protéine Kinase A (en bleu clair)

d'après Facchin et al 2002 Hecker et al. 2007, Mao et al. 2008,

Une autre particularité est le caractère « acidophile » de cette kinase (ce qui n'a rien à voir avec les organismes acidophiles !). Alors que la très grande majorité des Ser/Thr kinases reconnaissent les résidus phospho-accepteurs grâce à la présence d'un consensus de résidus basiques, prolyl et hydrophobes ; chez les rares Ser/thr kinases dites acidophiles, le choix du site de phosphorylation dépend de la présence de résidus acides, à chaîne latérale négative (comme la sérine) autour du résidu phosphoaccepteur (FACCHIN ET AL., 2002).

Au niveau structural, plusieurs caractéristiques communes aux kinases sont retrouvées chez Bud32 et ses orthologues. La comparaison de sa séquence avec celle de la PKA (Protéine Kinase AMPc dépendante) montre la présence conservée de certains motifs typiques des kinases, comme la boucle P, la boucle catalytique, et d'autres résidus conservés impliqués dans la fixation de l'ATP. Le rôle de ces résidus conservés dans l'activité kinase de Bud a été confirmé par étude mutationnelle sur la protéine de *S. cerevisiae* (FACCHIN ET AL., 2002, MAO ET AL., 2008) (FIGURE 48). La résolution de la structure de la protéine MJ1130 *Methanocaldococcus jannaschii*, où Bud32 est fusionnée avec Kae1 en un unique polypeptide, a permis de confirmer qu'elle présentait un repliement typique des kinases, avec deux lobes reliés par une charnière (FIGURE 48). Typiquement, le lobe N-terminal correspond au domaine de fixation du nucléotide, tandis que le lobe C-terminal est responsable de la fixation avec la protéine substrat pour effectuer le transfert du phosphate. Cette structure permet de rapprocher Bud32 de la famille des Rio kinases, impliquées dans la maturation du ribosome.

Cependant, sur de nombreux points, Bud32 diffère fortement des autres kinases et justifie son caractère « atypique ». D'une part, son domaine C-terminal est très réduit et ne possède pas de boucle A, normalement impliquée dans la reconnaissance du substrat. Il est terminé par une extrémité C-terminale basique atypique. Le site d'autophosphorylation, très réduit, ne présente pas les caractéristiques classiques pour ce type de kinase. Au niveau de son site actif se trouve une thréonine (conservée chez tous les orthologues) à la place d'une lysine caractéristique des Ser/Thr kinases (FACCHIN ET AL., 2002), cruciale pour le transfert du phosphate gamma.





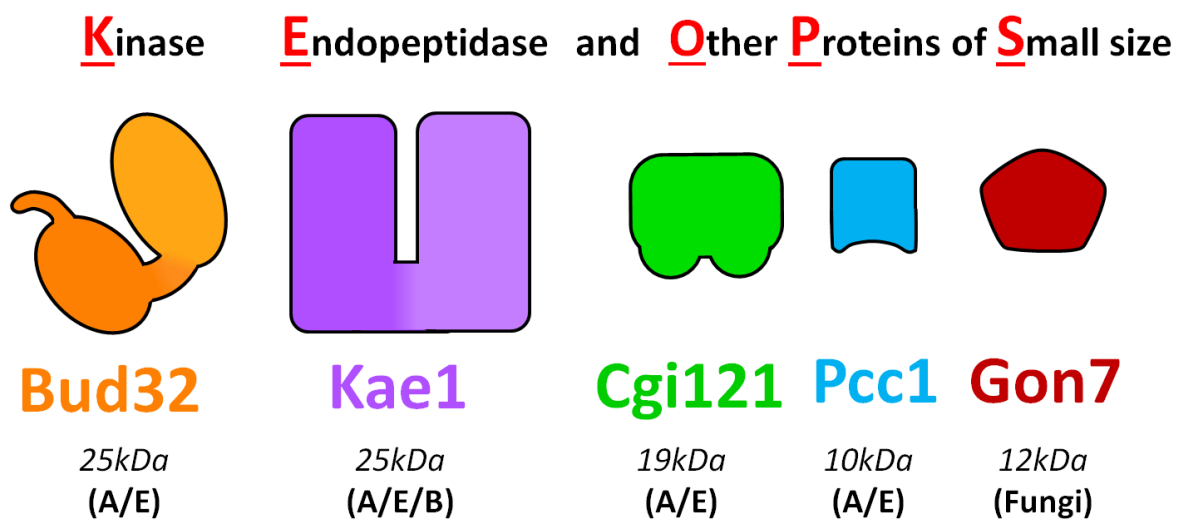
**Bud32 apparait donc comme une kinase atypique. Les données structurales et biochimiques suggèrent qu'elle fonctionne sur un mode de reconnaissance et un mécanisme catalytique différents des autres kinases, qui pourraient témoigner de l'état ancestral des kinases d'eucaryotes et d'archées.**

Au niveau *in vivo*, les mutants de délétion de Bud32 chez *S. cerevisiae* montrent, en plus d'une altération importante de la croissance, plusieurs défauts spécifiques dont : une survie réduite en phase stationnaire, une incapacité des homozygotes diploïdes à entrer en sporulation, aucune division méiotique visible, des altérations de la structure de la paroi, ainsi qu'un bourgeonnement (**budding**) aléatoire. C'est ce dernier phénotype qui donne son nom à la protéine (SNYDER ET AL 2001).

Chez l'Homme, l'homologue de Bud32 est appelé PRPK, pour p53 regulating protein kinase. En effet, il a été montré à la fois *in vivo* et *in vitro* que PRPK phosphoryle et active la protéine p53, régulateur transcriptionnel central du cycle cellulaire chez l'Homme (ABE ET AL., 2001). Plusieurs différences apparaissent entre PRPK et son homologue Bud32 (du magnésium est requis pour l'activité de PRPK, et il faut « l'activer » avec du lysat cellulaire pour obtenir une activité kinase), mais de manière intéressante, il a été montré que Bud32 exprimé dans les cellules humaines pouvait se lier à p53 et le phosphoryler. De plus, PRPK peut partiellement compléter la délétion de Bud32 chez la Levure (LOPREIATO ET AL., 2004). Ces données indiquent que la fonction de ces orthologues est conservée.

Récemment, il a été montré chez l'Archée *Haloferax volcanii* que la délétion du gène codant l'orthologue de Bud32 est létale, indiquant son caractère essentiel (NAOR ET AL., 2012)

Ainsi, la fonction de Bud32 semble toucher un processus cellulaire fondamental. Cependant, à part p53 (qui n'est pas présent chez *S. cerevisiae* ni chez les Archées), aucun substrat spécifique n'a pour le moment été identifié. De plus, la nécessité de « pré-activer » PRPK à l'aide de lysat cellulaire, et la faible activité kinase de Bud32 de Levure *in vitro*, permet de suggérer l'existence d'activateurs non identifiés. La recherche des partenaires et de substrats de Bud32/PRPK fut donc entreprise pour déterminer les processus cellulaires auxquels elle était liée. En plus de p53, plusieurs protéines interagissant avec Bud32/PRPK ont été identifiées chez les Eucaryotes: une protéine de fonction inconnue appelée Cgi121



**FIGURE 49: Composition et nomenclature du complexe KEOPS**

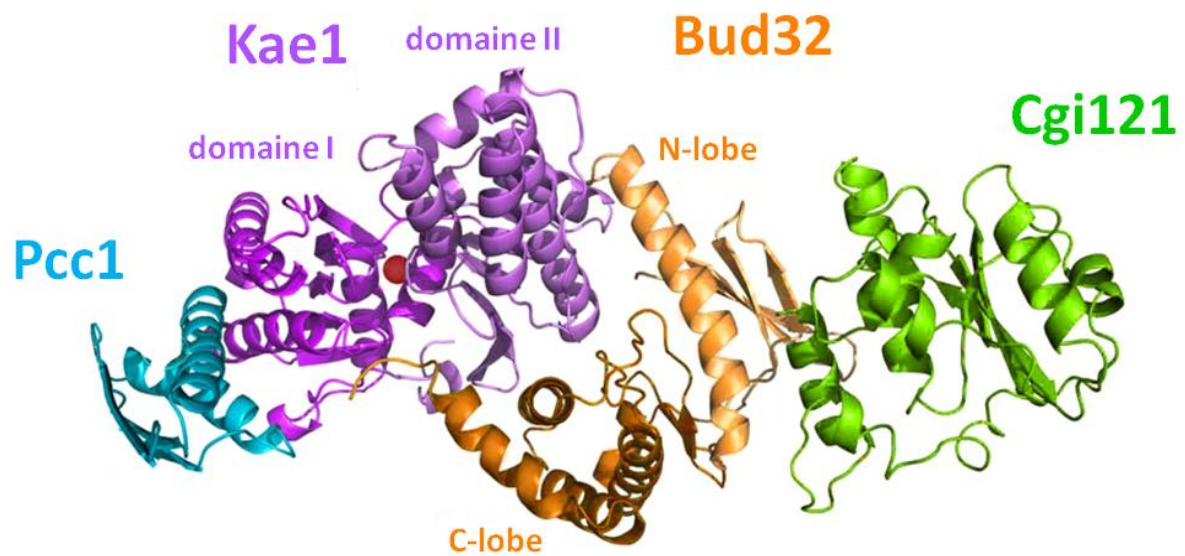
Les noms et les masses moléculaires indiquées correspondent aux protéines de *Saccharomyces cerevisiae*. Les formes de chaque protéine sont des représentations schématiques simplifiées de la structure résolue chez les Archées (d'après Mao et al, 2008), et les couleurs ont été définies arbitrairement. Entre parenthèse sont indiqués les groupes d'organismes où sont retrouvés des orthologues (A= Archées, E= Eucaryotes, B= Bactéries). La protéine Gon7 n'est retrouvée que chez les Champignons (Fungi) dont *S. cerevisiae*.

(MYIOSHI ET AL., 2003), une glutaredoxine Grx4, la protéine Kae1 (LOPREIATO ET AL., 2004), Ray (une Ras-like GTPase) (ABE ET AL., 2006), AKT, une protéine centrale de la signalisation cellulaire (PINNA ET AL., 2007). Ces interactants suggèrent que Bud32/PRPK est au cœur d'un réseau complexe de signalisation, du moins chez les Eucaryotes.

Parmi cette liste de partenaires identifiés, deux protéines se distinguent pour différentes raisons. D'une part, la protéine Cgi121 (dont le nom correspond uniquement à un nom de code issu d'une banque d'Expressed Sequenced Tag, ou EST), présente des orthologues chez les autres Eucaryotes et chez les Archées, mais pas chez les Bactéries, ce qui correspond à la distribution de Bud32 dans le monde vivant. D'autre part, la protéine Kae1, protéine universelle, est parfois fusionnée à Bud32 chez certaines Archées comme chez *Methanocaldococcus jannaschii* (HUNTER ET AL., 1997, KOONIN ET AL., 1998), indiquant un lien physique et fonctionnel conservé.

En 2006, deux études parallèles chez *S. cerevisiae* confirmèrent l'interaction entre Bud32, Kae1 et Cgi121 et complétèrent la liste des interactants. La première étude (DOWNEY ET AL. 2006) montra que ces trois protéines forment dans la cellule un complexe stable et isolable, contenant également une petite protéine spécifique des Champignons appelée Gon7. Ce complexe fut appelé **KEOPS** pour Kinase (correspondant à Bud32), Endopeptidase (correspondant à l'activité supposée de Kae1) and Other Proteins of small Size (désignant les petites protéines Cgi121 et Gon7). La seconde étude (KISSELEVA ET AL., 2006), identifia un complexe similaire, mais avec en plus une autre petite protéine : Pcc1. Cette dernière, tout comme Bud32 et Cgi121, est universellement conservée chez les Eucaryotes et les Archées, mais absente chez les Bactéries. Ce complexes fut appelé quant à lui **EKC**, signifiant Endopeptidase (correspondant à Kae1), Kinase (correspondant à Bud32) Chromatine-associated (car ce complexe est retrouvé associé à la chromatine lors d'expérience d'immunoprécipitation) L'association entre les protéines Kae1, Bud32, Cgi121 et Pcc1 fut par la suite confirmée chez les Archées (MAO ET AL., 2008), puis chez l'Homme (COSTESSI ET AL., 2012).

Dans la suite de ce manuscrit, seule l'appellation **KEOPS** sera employée, et désignera le complexe stable constitué des orthologues des 4 protéines : Kae1, Bud32, Cgi121, Pcc1 (auxquelles s'ajoute Gon7 dans le cas de la Levure) (**FIGURE 49**)



**FIGURE 50:** Architecture linéaire du complexe KEOPS chez les Archées

Modèle élaboré à partir d'études interactionnelles et de la résolution de la structure de différents sous-complexes D'après Mao *et al*, 2008

## 2 - Propriétés structurales et biochimiques du complexe KEOPS

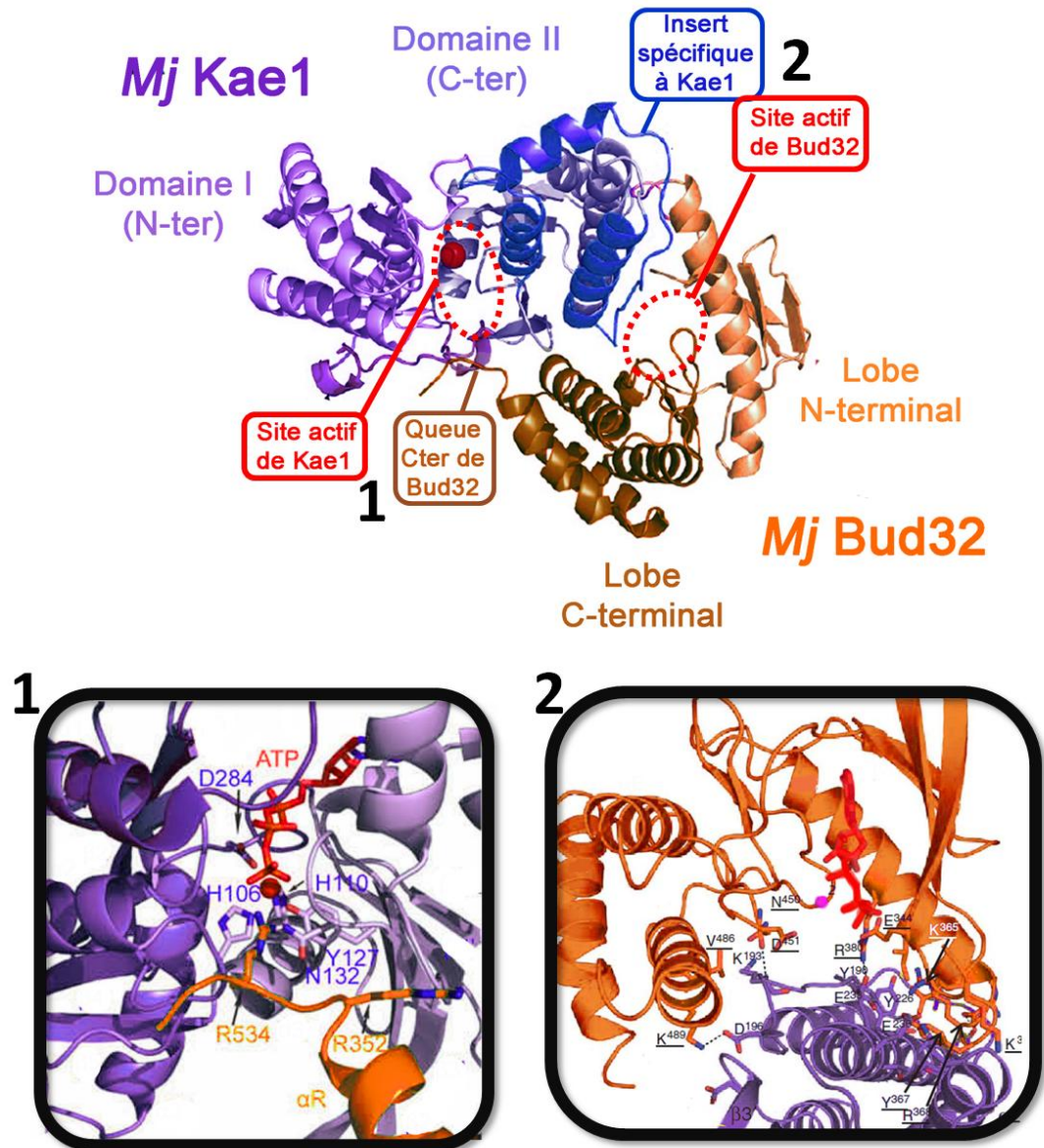
### a) Une architecture linéaire :

Lors de son identification, la masse moléculaire apparente du complexe KEOPS chez *S. cerevisiae* avait été estimée entre 300 et 400kDa par gel filtration (DOWNEY ET AL., 2006, KISSELEVA ET AL., 2006). Or, en considérant un complexe comportant un seul exemplaire de chaque protéine, la masse moléculaire attendue est de l'ordre de 117kDa. Cela suggère une stoechiométrie et/ou une forme moléculaire non globulaire. Les analyses d'interaction *in vitro* des différentes protéines d'Archées montrent que le complexe KEOPS présente une architecture linéaire, formée par l'enchaînement des protéines Pcc1 – Kae1 – Bud32 – Cgi121 (MAO ET AL., 2008) (FIGURE 50). De plus, les données structurales suggèrent que le complexe pourrait former un homodimère, dont la masse moléculaire théorique est de 233kDa. Ainsi, ces données permettraient d'expliquer la masse moléculaire apparente très élevée déterminée lors de l'isolation initiale du complexe KEOPS chez *S. cerevisiae*. L'architecture linéaire et l'état dimérique ont également été décrits pour le complexe KEOPS humain (COSTESSI ET AL., 2012).

### b) Kae1/Bud32 : un couple très uni

L'interaction entre la protéine Kae1 et Bud32, qui constituent la partie centrale du complexe, a été caractérisée grâce à la résolution de la structure de la protéine MJ1130 de *M. janaaschii*, où les deux protéines sont fusionnées en un seul polypeptide (HECKER ET AL., 2008, MAO ET AL., 2008) (FIGURE 51). Combinées à une caractérisation biochimiques *in vitro* ainsi qu'à une analyse mutationnelle *in vivo* des résidus conservés chez les orthologues de *S. cerevisiae*, ces deux études ont permis de déterminer le rôle et l'importance de certaines régions dans l'interaction entre ces deux protéines.

Bud32 interagit avec Kae1 via ses deux lobes N et C terminaux, alors que du côté de Kae1, l'interaction engage principalement le domaine II. La région appelée insert spécifique 1 de ce domaine joue le rôle d'une plate-forme d'interaction avec Bud32. De manière intéressante, la zone de contact entre Bud32 de cet insert n'est conservée que chez les orthologues de Kae1 d'Archées et d'Eucaryotes, mais absent chez les orthologues Bactériens



**FIGURE 51 :** Structure de la protéine MJ1130 de *M.jannaschii*, correspondant à une fusion entre Kae1 et Bud32, et mise en évidence des interactions entre ces deux domaines.

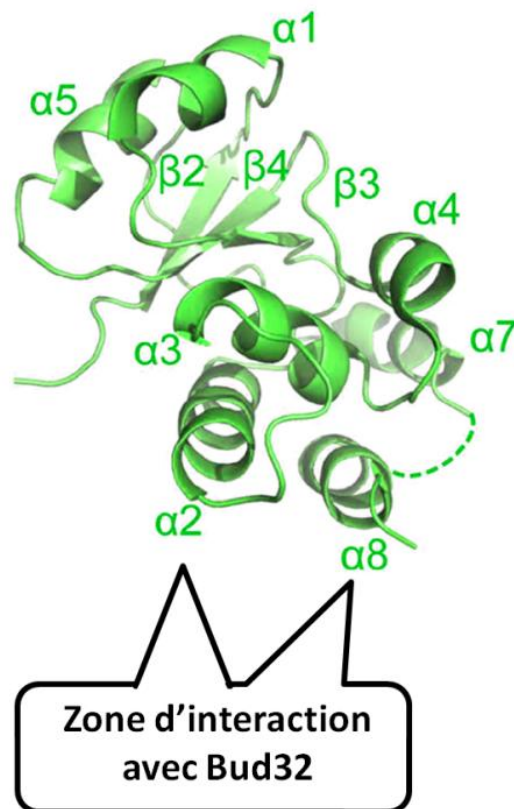
Le domaine Kae1 est coloré en violet et en bleu, le domaine Bud32 en orange. Les différents éléments caractéristiques sont indiqués. L'atome de Fer est coloré en rouge. Les encarts 1 et 2 correspondent à un agrandissement du site actif des deux domaines. Dans l'encart 1 (Kae1), une molécule d'ATP a été modélisée (en rouge) d'après la structure de Kae1 de *P. abyssi*. On voit la queue Cterminal de Bud32 qui se projet au sein du site actif (en orange). Dans l'encart 2, une molécule d'ATP (en rouge) et un atome de manganèse (en rose) ont été modélisés au sein du site actif. Les résidus conservés et impliqués dans la coordination des ligands et des interactions entre protéines sont indiqués. d'après Mao *et al*, 2008.

(HECKER ET AL., 2008). Par comparaison avec la structure de Kae1 de *Pyrococcus abyssi* (HECKER ET AL., 2007), il semble que la présence de Bud32 ne modifie pas la conformation globale de Kae1. L'effet de Kae1 sur la conformation de Bud32 n'est en revanche pas connu, car aucune structure de Bud32 seule n'a pour le moment été décrite. Certains résidus chez Kae1 impliqués dans cette interaction sont très proches du site actif de Bud32, et forment mêmes des liaisons hydrogènes avec certaines bases catalytiques (comme l'aspartate conservé 451). Dans les différentes structures résolues, aucun nucléotide n'a pu être co-cristallisé au sein du site actif de Bud32, et il semble que sa *P-loop* soit dans une conformation inactive (ou fermée). De plus, Hecker *et al.* ont montré que la présence de Kae1 inhibe l'activité d'autophosphorylation de Bud32 *in vitro*. L'ensemble de ces données suggèrent que Kae1 pourrait réprimer l'activité kinase de Bud32 par effet conformationnel sur son site actif. A l'inverse, Mao et al. ont montré que Kae1 est la cible de phosphorylation de Bud32. Bien que contradictoires au niveau biochimique, ces deux études ont montré l'importance *in vivo* de cette surface d'interaction : la substitution des résidus au niveau de l'interface affecte fortement la croissance chez la Levure.

Enfin, une autre région d'interaction est également particulièrement intéressante. Les structures révèlent que la queue C-terminale de Bud32 pénètre au sein de la fente catalytique de Kae1, directement à côté du centre de coordination du Fer. La délétion de cette région chez Bud32 ne perturbe pas l'interaction avec Kae1, mais affecte fortement la croissance chez la Levure, indiquant que cette région est nécessaire à la fonction de Bud32 ou du couple Bud32/Kae1 (MAO ET AL., 2008)

<p><b>Au bilan, il apparaît que Kae1 et Bud32 interagissent physiquement et fonctionnellement de façon réciproque l'une sur l'autre, par l'intermédiaire de zones d'interaction situées à proximité de leurs sites actifs respectifs. Cette interaction mutuelle joue probablement un rôle dans la régulation de leurs activités respectives et dans la fonction biologique de ce complexe protéique</b></p>
--





**FIGURE 52** : Structure de la protéine Cgi121 de *M.jannaschii*.

Les deux hélices  $\alpha 2$  et 8 impliquées dans l'interaction avec Bud32 sont indiquées.



### **c) Cgi121, un régulateur de Bud32**

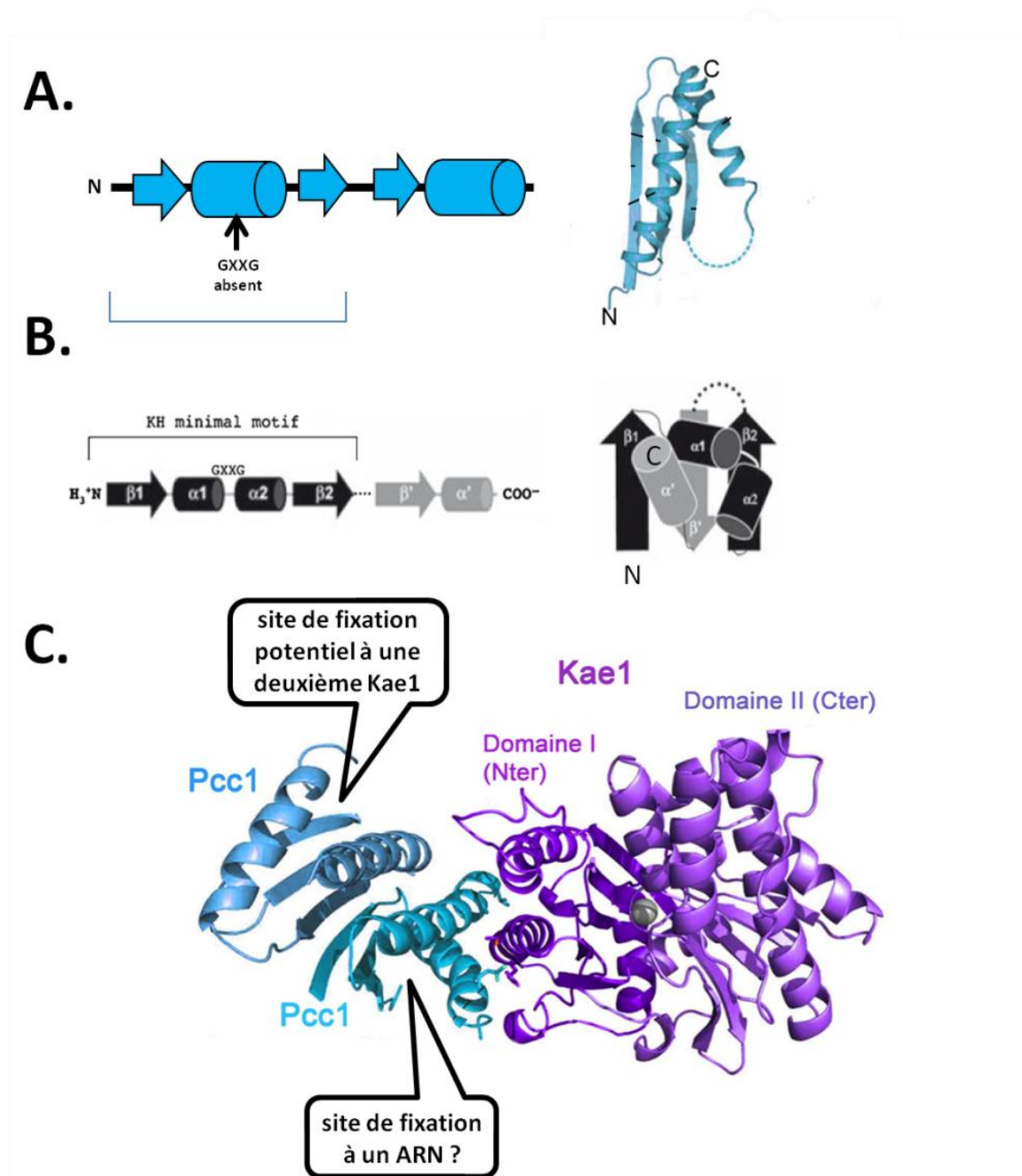
L'étude de Mao et al. a également révélé la structure de Cgi121 chez l'Homme et chez l'Archée *M.jannaschii* (MAO ET AL., 2008). Ces deux orthologues présentent une conformation très proche, mais qui ne correspond à aucun domaine structural connu. La structure de Cgi121 d'Archée a également été obtenue au sein du complexe Kae1-Bud32-Cgi121, permettant de révéler son mode d'interaction avec Bud32. Cgi121 interagit via deux hélices alpha (1 et 8), au niveau du lobe N-terminal de Bud32 (FIGURE 52). Cette interaction provoque des changements conformationnels notables dans la structure de ce lobe.

*In vitro*, la présence de Cgi121 augmente l'activité d'autophosphorylation de Bud32. Des mutations dans la zone d'interaction de Cgi121 ou sa délétion, provoquent un phénotype de croissance semblable à celui provoqué par un mutant catalytique de Bud32. De plus, chez la Levure, Cgi121 semble être moins exprimée que les autres membres de KEOPS et soumise à une dégradation plus rapide (MAO ET AL., 2008, GHAEMMAGHAMI ET AL., 2003). L'ensemble de ces données structurales et biochimiques rappellent l'action des cyclines au niveau des Cycline dépendant kinases (CDK), impliquées dans la régulation du cycle cellulaire chez les Eucaryotes.

**Au bilan, Cgi121 apparait comme un probable régulateur positif de Bud32.**

### **d) Pcc1, partenaire de Kae1**

Pcc1 (pour *Polarized growth Chromatin associated controller 1*, KISSELEVA ET AL., 2006) est la plus petite des 4 protéines du complexe KEOPS. Sa structure, résolue pour l'Archée *Pyrococcus furiosus* (MAO ET AL., 2008) montre une certaine ressemblance avec le domaine conservé appelé KH (FIGURE 53 A-B). Le domaine KH (d'après la protéine hnRNP K chez l'Homme), constitué généralement d'environ 70 acides aminés, est retrouvé dans les trois domaines du vivant, au sein de nombreuses protéines, notamment de nombreux régulateurs transcriptionnels. C'est un domaine de fixation à l'ARN ou ADN simple brin qui accommode strictement 4 bases de l'acide nucléique au sein d'une cavité hydrophobe (VALVERDE ET AL., 2008). Les domaines KH présentent tous un motif conservé GXXG impliqué dans l'interaction avec l'acide nucléique, bien qu'il ait été montré récemment que des domaines KH divergents, ayant un motif GXXG altéré, peuvent être fonctionnels (BRYKAILO ET AL., 2007). Enfin, ce domaine cristallise très souvent sous forme de dimère et pourrait avoir un rôle dans la dimérisation des protéines qui présentent ce domaine (VALVERDE ET AL., 2008).



**FIGURE 53 :** Structure de la protéine Pcc1 de *T. acidophilum*, comparaison avec le domaine KH et structure du complexe Kae1/Pcc1.

A. Représentation schématique de la structure secondaire de Pcc1. Les cylindres représentent des hélices  $\alpha$ , les flèches des brins  $\beta$ . L'indication GXXG indique la position théorique du motif conservé dans les domaines KH, qui n'est pas présent dans la séquence de Pcc1. Sur la droite est représentée la structure d'un monomère de Pcc1 (obtenu sous forme de dimère)

B. Représentation schématique de la structure secondaire et tertiaire d'un domaine KH. Les hélices  $\alpha 1$  et  $\alpha 2$  correspondent à une unique hélice interrompue par un coude correspondant au motif conservé GXXG. Les éléments en noirs correspondent aux structures retrouvées systématiquement dans tous les domaines KH.

C. Structure du dimère de Pcc1 (en bleu) avec un monomère de Kae1 (en violet).

d'après Mao *et al.* 2008, Valverde *et al.*, 2008

Pcc1 présente les éléments structuraux typiques de ce domaine KH, hormis le motif GXXG dans sa première hélice. Il dimérise en solution et cristallise également sous forme de dimère, selon un mode semblable aux domaines KH (MAO ET AL., 2008).

De manière intéressante, Pcc1 interagit avec Kae1 en solution selon une stoechiométrie 2:2, c'est-à-dire sous forme d'un homodimère Kae1Pcc1-Pcc1Kae1 et des analyses mutationnelles chez la Levure semblent indiquer que l'ensemble du complexe KEOPS pourrait fonctionner *in vivo* comme un homodimère via Pcc1 (MAO ET AL., 2008). Chez l'Homme (COSTESSI ET AL., 2012), un homodimère semblable : Kae1Pcc1-Pcc1Kae1 (Pcc1 est appelé ici LAGE3) a été décrit.

Au niveau structural, seule la structure d'un dimère de Pcc1 (de *Pyrococcus furiosus*) lié à une seule molécule de Kae1 (de *Thermoplasma acidophilum*) a pu être obtenue (FIGURE 53-C) (MAO ET AL., 2008). Pcc1 interagit au niveau du domaine I de Kae1, et ne semble pas provoquer pas de réarrangement conformationnel notable dans la structure de cette dernière, par comparaison avec la structure de la protéine Kae1 de *P. abyssi* seule. (HECKER ET AL. 2007)

**Ainsi, Pcc1 pourrait avoir une double fonction au sein de KEOPS : dimérisation et liaison avec un acide nucléique par sa ressemblance au domaine KH. Ces hypothèses nécessitent toutefois une confirmation.**

**La protéine Kae1 est donc associée chez les Archées et les Eucaryotes à une kinase atypique, Bud32, ainsi qu'à deux autres protéines, Pcc1 et Cgi121, pour former un complexe moléculaire conservé, KEOPS. L'étude structurale de plusieurs sous-complexes de KEOPS d'Archées a permis de proposer un modèle probable d'architecture globale, et a fourni de nombreuses données sur les relations intra et inter-moléculaires. Chacune des protéines semble influencer l'activité ou la structure de son ou ses partenaires respectifs. Cependant, ces données ne permettent pas d'établir une fonction biologique pour ce complexe**



### 3 - KEOPS : un rôle biologique mystérieux mais fondamental:

Les différentes sous-unités de KEOPS ont été l'objet de multiples investigations génétiques afin de déterminer leur fonction. Cgi121 fut identifié initialement chez *S. cerevisiae* comme suppresseur d'une mutation dans le gène *cdc13* qui code pour une protéine impliquée dans l'homéostasie des télomères. La délétion de *cgi121*, *bud32* et *gon7* (le cinquième composant de KEOPS chez la Levure) provoque un raccourcissement des télomères associé à un ralentissement de la croissance, phénotype semblable à celui associé à la délétion du gène *sua5*. De manière intéressante, la délétion de *bud32* engendre un phénotype plus sévère que la délétion de *cgi121* (DOWNEY ET AL., 2006). D'autre part, la délétion de *pcc1* réduit fortement la croissance cellulaire et est associé un défaut de transcription de différents gènes, notamment les gènes de réponse aux phéromones et les gènes activés par le galactose (GAL) (KISSELEVA ET AL., 2006). Par immuno-précipitation de chromatine, il a été montré que Pcc1, Kae1 et Gon7, sont physiquement associés à certaines zones activement transcrites de la chromatine. Finalement, l'étude de Kisseleva suggère que le complexe entier est impliqué dans la régulation de la transcription, en agissant sur un remodelage de la chromatine (d'où le nom initial de EKC, C pour *chromatine associated*). Enfin, chez l'Archée *Haloferax volcanii*, les gènes codant Kae1, Bud32, et Cgi121 sont essentiels, alors qu'une délétion de Pcc1 ne provoque pas la mort de l'organisme, mais entraîne un retard de croissance, associé à différents phénotypes, tels qu'une accumulation de l'ADN et une distribution anormale, une accumulation des produits de glycation avancés (des protéines altérées par glycation non enzymatique et traduisant des perturbations métaboliques) (NAOR ET AL., 2012).

Ces différentes observations, combinées aux résultats obtenus par les précédentes études concernant Bud32, semblent indiquer que les membres de KEOPS participent à de nombreux processus cellulaires fondamentaux. La pléiotropie des phénotypes observés ne permet pas d'établir un rôle biologique clair pour le complexe KEOPS, mais rappelle fortement les observations faites pour la famille YgjD/Kae1 et YrdC/Sua5



#### 4 - KEOPS, impliqué dans la biosynthèse de t<sup>6</sup>A ?

Kae1 étant intimement liée à Pcc1, Bud32 et Cgi121 au sein du complexe KEOPS, on peut donc se demander si ces différentes sous-unités participent également à la biosynthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes, aux côtés de Sua5. Différents travaux récents de génétique ont tentés de répondre à cette question.

**SRINIVASAN ET AL.** ont employé une technique d'extension d'amorces et une quantification par spectrométrie de masse sur des ARNt extraits de différents mutants de KEOPS afin de déterminer l'importance de chacune des ces protéines dans la formation de t<sup>6</sup>A *in vivo*. D'après les auteurs, les résultats obtenus indiquent que les protéines Sua5, Kae1, mais aussi Bud32 et Pcc1 sont nécessaires à la synthèse de t<sup>6</sup>A. En revanche, la protéine Cgi121 ne semble pas nécessaire. Cependant, la méthode d'inactivation des gènes employés peut être remise en question (communication personnelle, Valérie de Crécy-lagard), de même que les résultats concernant l'extension d'amorces qui ne correspondent pas aux profils attendus par cette technique. D'autre part, **DAUGERON ET AL** ont montré que des mutants ponctuels thermosensibles de Kae1 et Pcc1 chez *S. cerevisiae* présentent une diminution de la quantité de t<sup>6</sup>A dans les conditions non permissives. Chez l'Archée *Haloferax volcanii*, l'absence de Pcc1 conduit également à une légère diminution de la quantité de t<sup>6</sup>A au niveau des ARNt, suggérant que cette protéine pourrait jouer un rôle accessoire dans la synthèse de ce nucléoside (**NAOR ET AL., 2012**)

Enfin, une première approche de biosynthèse de t<sup>6</sup>A *in vitro* a été tentée par **SRINIVASAN ET AL** en mesurant la quantité de thréonine radioactive(C14) ou lourde (N13) dans des ARNt en présence d'extraits cellulaires issus des mutants de KEOPS. Les résultats obtenus, très préliminaires et en partie contradictoires, semblent indiquer que KEOPS est impliqué directement dans cette synthèse.

**L'ensemble de ces résultats suggèrent que les membres de KEOPS sont effectivement impliqués dans la voie de biosynthèse de t<sup>6</sup>A. Cependant, les données disponibles ne permettent pas d'établir des conclusions définitives quant au rôle de ces protéines dans cette biosynthèse. Il manque donc la démonstration biochimique *in vitro* prouvant que KEOPS et Sua5 constituent, à l'image des protéines YrdC, YgjD, YeaZ et YjeE chez les Bactéries, la machinerie de biosynthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes. C'est la question principale à laquelle ce travail de thèse se propose de répondre.**





# OBJECTIFS

## PROBLEMATIQUE

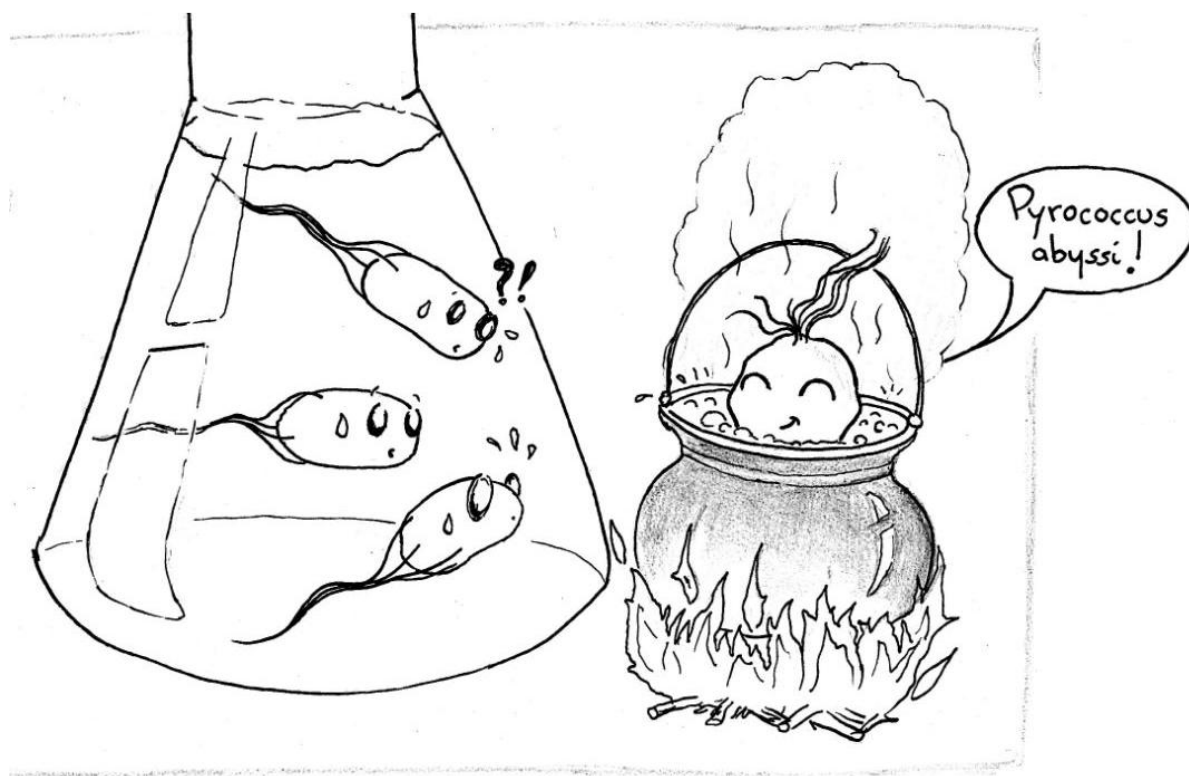
Basée sur les données présentées dans l'introduction, cette thèse a pour objectif de répondre à différentes questions :

- Chez les Archées et les Eucaryotes, la protéine Sua5 et le complexe KEOPS forment-ils la machinerie de biosynthèse de t<sup>6</sup>A ?
- Quel est le rôle des différents composants de cette machinerie ?
- Chez la mitochondrie, comment est synthétisée t<sup>6</sup>A ?
- Quels sont les différences et les points communs entre ces différents types de machinerie enzymatique ? Peut-on retracer leur histoire évolutive ?

## CHOIX DE L'APPROCHE

Afin de répondre à ces questions, je me suis orienté sur une approche biochimique *in vitro*, par l'étude, la caractérisation et la comparaison de protéines issues de différents organismes modèles. D'une part, aucune étude de cette nature concernant le rôle de Sua5 et KEOPS dans la biosynthèse de t<sup>6</sup>A n'a été décrite pour le moment chez les Archées et les Eucaryotes. D'autre part, cette approche a l'avantage, contrairement aux techniques de génétique, de démontrer l'implication directe des protéines candidates. Cependant, elle ne reflète pas nécessairement les conditions physiologiques et cellulaires naturelles (en terme de concentration, de présence de co-facteurs) et doit donc être interprétée en regard des données disponibles obtenues par d'autres techniques.

Les protéines qui font l'objet de cette étude ont toutes été produites en système hétérologue, par sur-expression dans des souches adaptées d'*Escherichia coli*. Ce mode de production a prouvé son efficacité et sa fiabilité, bien que l'obtention de protéines solubles et actives par cette méthode soit empirique et non prévisible. De plus, les protéines obtenues ne portent pas les modifications post-traductionnelles éventuelles que l'on pourrait rencontrer *in vivo* (phosphorylation), et qui pourraient agir sur leur activité.



DESSIN ADDITIONNEL :  
***Pyrococcus abyssi*, un modèle d'étude... original !**

Enfin, pour faciliter les étapes de purification, chaque protéine est fusionnée avec une étiquette peptidique (hexahistidine ou StrepTagII) placée généralement du côté C-terminale. En raison de sa taille (6 et 8 acides aminés respectivement) nous avons choisi de ne pas l'enlever. La présence d'une telle étiquette peut dans certains cas perturber le repliement, l'activité ou l'interaction d'une protéine avec ses partenaires, et il faut donc garder à l'esprit l'existence de ce paramètre dans l'interprétation des données.

## **CHOIX DES MODELES D'ETUDES**

Le modèle d'étude principal choisi ici sera l'Euryarchée hyperthermophile *Pyrococcus abyssi*. Ce choix repose principalement sur des raisons historiques : découverte en 1992 dans l'océan Pacifique, au niveau d'un fumeur noir, son génome a été séquencé et caractérisé en 2005 par le laboratoire d'Orsay. Il a servi ainsi de modèle à de nombreuses études sur la biologie des Archées. De plus, *Pyrococcus abyssi* a l'avantage d'être très proche de l'Archée *Thermococcus kodakarensis*, pour laquelle des outils génétiques sont disponibles. Enfin, au niveau biochimique, l'étude des protéines issues d'un organisme hyperthermophile présentent souvent des avantages (purification, meilleure stabilité, étude structurale facilitée, températures de réaction garantissant l'absence d'activités contaminantes). *Pyrococcus abyssi* présente un orthologue identifié de chacune des protéines Sua5, Kae1, Bud32, Cgi121, et Pcc1.

*Saccharomyces cerevisiae* sera le second modèle utilisé dans cette étude. Le complexe KEOPS a été caractérisé pour la première fois chez cet Eucaryote modèle, et de très nombreuses données génétiques sont disponibles concernant les différentes protéines décrites précédemment. Cependant, le complexe KEOPS de la Levure présente la particularité de contenir un cinquième élément, la protéine Gon7, spécifique des Champignons. Ainsi, la composition du complexe KEOPS de *Pyrococcus abyssi* (avec uniquement 4 sous unités) est plus représentative pour l'établissement d'un modèle général chez les Eucaryotes et les Archées, illustrant le fait que dans de très nombreux cas, les mécanismes archéens représentent un excellent modèle simplifié des mécanismes eucaryotes. La protéine Qri7, orthologue de Kae1/YgjD et présente dans la mitochondrie de *S. cerevisiae*, sera quant à elle utilisée comme modèle pour la compréhension de la machinerie mitochondriale.



Enfin, les protéines YrdC, YgjD, YjeE et YeaZ de la Bactérie modèle *Escherichia coli* seront utilisées dans l'optique de comparer les systèmes de biosynthèse de t<sup>6</sup>A issus d'organismes représentatifs des trois domaines du vivant.

## **DESCRIPTION DU CONTENU:**

Dans une première partie (**Partie I**), je répondrai à la question suivante : « Sua5 et KEOPS constituent-ils la machinerie de synthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes ? ». Après une brève présentation des méthodes employées et des principaux résultats obtenus, cette partie sera présentée sous la forme d'un article publié en Janvier 2013 dans le journal Nucleic Acid Research. Des résultats complémentaires seront présentés et discutés par la suite en addition des résultats décrits dans cette publication.

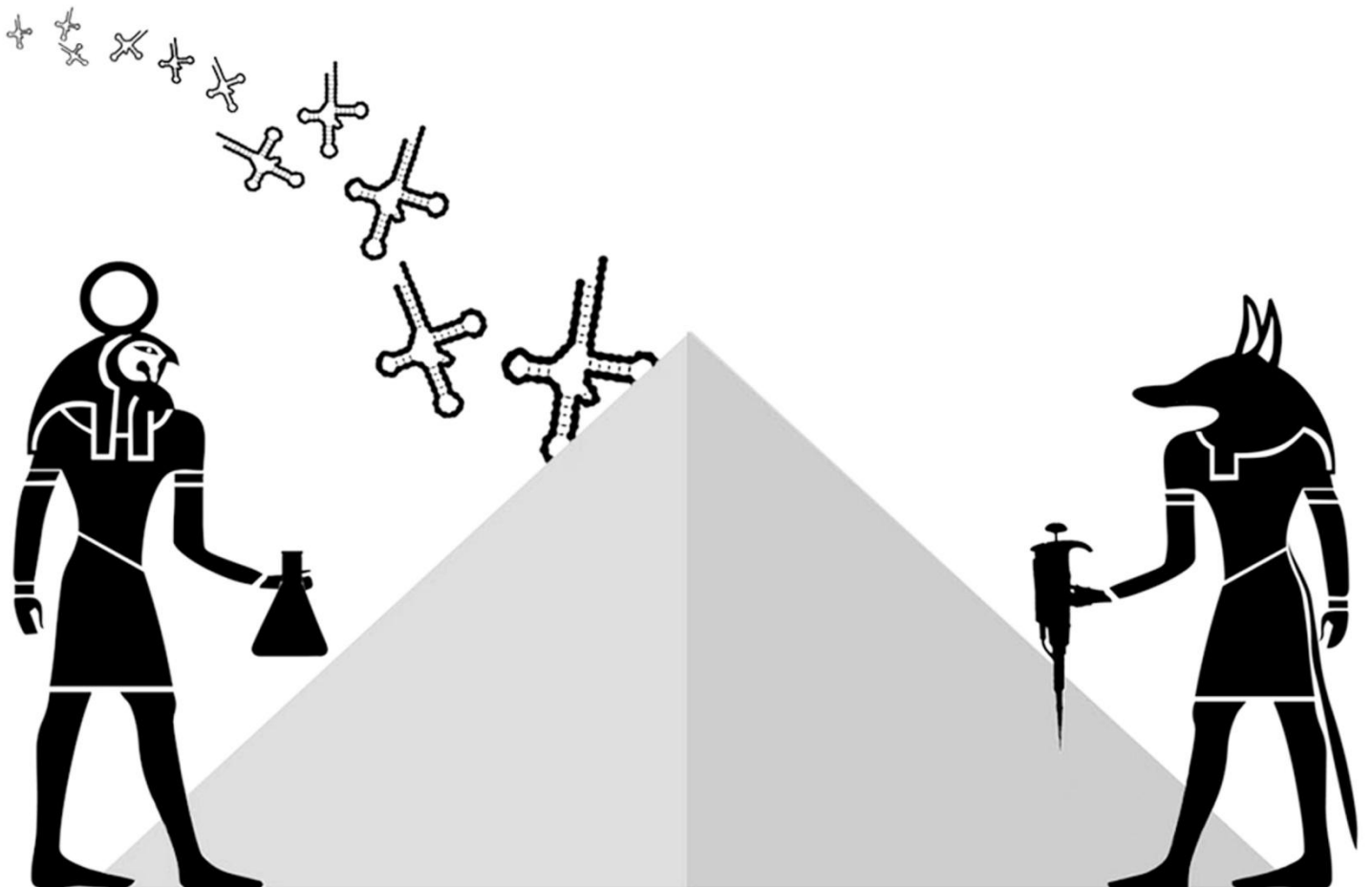
Après avoir montré l'implication du complexe KEOPS dans la voie de biosynthèse de t<sup>6</sup>A, je tenterai dans une seconde partie (**Partie II**) de répondre à la question « Quels sont les rôles respectifs des différents composants du complexe KEOPS ? », en caractérisant les propriétés biochimiques des sous-complexes de KEOPS. Les résultats obtenus permettront de proposer, en regard des données disponibles dans la littérature, un modèle de fonctionnement global du complexe KEOPS.

Enfin, dans une dernière partie (**Partie III**), j'aborderai la question de la synthèse de t<sup>6</sup>A chez la mitochondrie, en montrant que le système enzymatique nécessite uniquement deux composants et constitue de ce fait une version minimaliste de la machinerie de biosynthèse de t<sup>6</sup>A.

Je conclurai par la comparaison des différentes machineries de synthèse de t<sup>6</sup>A, les grandes questions qui en découlent, et les scénarios évolutifs possibles.



# RESULTATS ET DISCUSSION







# **PARTIE I : Sua5 et KEOPS constituent le système enzymatique de biosynthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes**



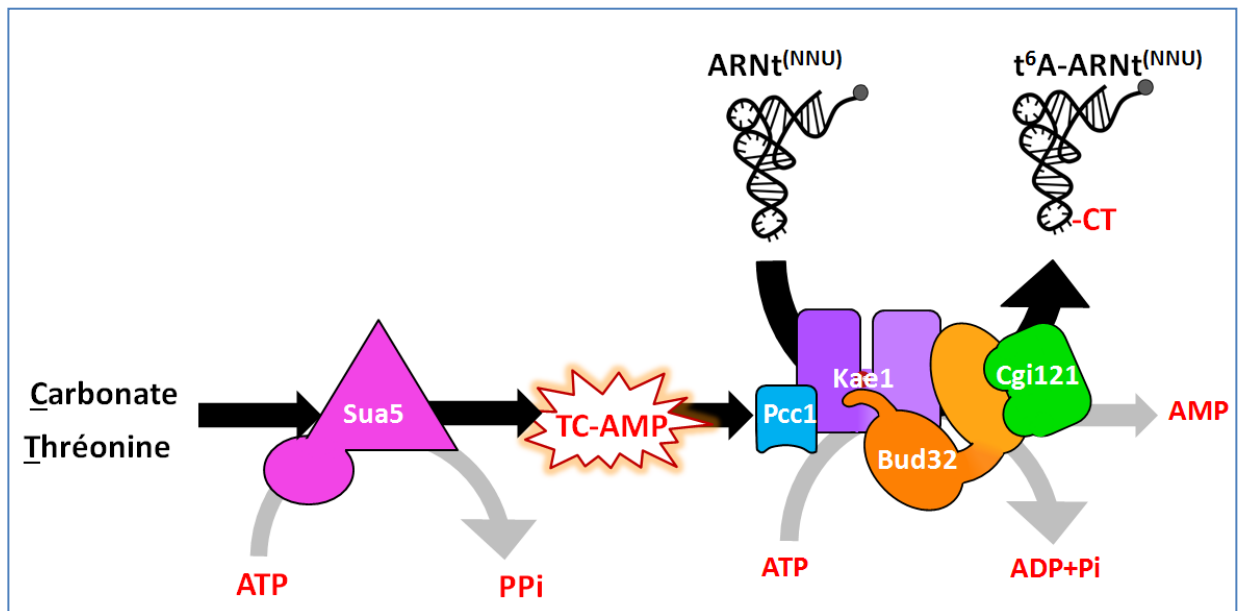
## Résumé des résultats principaux

La première question à laquelle nous avons tenté de répondre est donc la suivante: quelles sont les protéines qui catalysent la synthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes ?

Grâce aux précédents travaux réalisés au sein de notre équipe, nous disposons de la protéine Kae1 de *Pyrococcus abyssi* purifiée sous la forme d'une protéine recombinante portant une étiquette hexahistidine. Nous avons purifié de la même manière la protéine Sua5 de *Pyrococcus abyssi*. Enfin, nous avons produit le complexe KEOPS en co-exprimant simultanément les protéines Kae1, Bud32, Cgi121 et Pcc1 de *P.abys*si au sein d'une souche d'*Escherichia coli*. La présence d'une étiquette hexahistidine au niveau de la protéine Pcc1 permet de purifier ces quatre protéines sous la forme d'un complexe protéique de taille apparente proche de 200kDa. La taille théorique du complexe KEOPS étant de 90KDa, cela suggère que le complexe obtenu est sous forme dimérique en solution.

De manière intéressante, ce complexe co-purifie avec certains ARNt issus de la souche d'expression. La protéine Sua5, en revanche, ne présente pas cette particularité. Nous avons alors confirmé l'affinité du complexe KEOPS pour les acides nucléiques par retard sur gel. KEOPS fixe de manière non spécifique un ARN 65mer de séquence aléatoire aussi bien qu'un ARNt (ANN ou non ANN) dans nos conditions *in vitro*. Cela suggère que la reconnaissance des ARNt par le complexe KEOPS intervient après la fixation. La protéine Sua5, en revanche, ne présente aucune affinité pour les acides nucléiques.

Après avoir obtenu ces différentes protéines, nous avons produit différentes espèces d'ARNt pouvant servir de substrat à la réaction. Nous avons alors suivi l'incorporation de L-thréonine radiomarkée au Carbone 14 au sein des ARNt, après incubation en présence des protéines et différents co-facteurs. Nous avons montré que les deux protéines universelles Kae1 et Sua5 à elles seules ne permettent pas d'incorporer la thréonine. En revanche, la protéine de Sua5 et le complexe KEOPS catalysent ensemble l'incorporation de thréonine au niveau de l'ARNt. A l'issue de cette réaction, l'analyse des nucléosides issus de la digestion des ARNt par chromatographie sur couche mince nous a permis de démontrer que l'incorporation de thréonine observée correspond effectivement à la synthèse de t<sup>6</sup>A. Cette réaction s'avère dépendante de la présence d'ATP, et la présence de manganèse dans le milieu réactionnel stimule cette réaction. Le rendement maximum de réaction que nous



**Figure R1=** Modèle globale de biosynthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes

TC-AMP= thréonylcarbamoyl-AMP, CT= N6thréonylcarbamoyl greffée sur l'adénosine 37 de l'ARNt.

ARNt(NNU) = ARN de transfert portant un anticodon de type NNU, où N= A, U, G ou C

avons pu obtenir est d'environ 50% d'ARNt modifié, et la synthèse de t<sup>6</sup>A n'a lieu uniquement sur les espèces d'ARNt décodant les codons ANN (Lys, Asn, Ile), conformément à la distribution *in vivo* de ce nucléoside modifié.

Cette réaction étant dépendante de la présence d'ATP, nous avons suivi par chromatographie sur couche mince l'apparition des produits d'hydrolyse d'ATP<sup>[αP<sup>32</sup>]</sup> radioactifs en présence de KEOPS et de Sua5. Nous avons montré que ces deux composants hydrolysent indépendamment l'ATP. La protéine Sua5 hydrolyse l'ATP en AMP en présence spécifique de thréonine et de carbonate. De manière intéressante, l'hydroxynorvaline, un acide aminé présent au sein d'une modification d'ARNt proche de t<sup>6</sup>A, hn<sup>6</sup>A, stimule également cette activité, suggérant que Sua5 pourrait être impliquée dans la formation de ce nucléoside modifié. L'activité d'hydrolyse de Sua5 en présence des précurseurs de t<sup>6</sup>A suggère la formation d'un intermédiaire réactionnel que nous n'avons pas isolé, mais qui a été identifié par d'autres études comme étant du TC-AMP (thréonyl-carbamoyl-AMP).

Le complexe KEOPS hydrolyse l'ATP en ADP. La présence d'un oligonucléotide ADN ou ARN de séquence aléatoire ne stimule pas cette activité d'hydrolyse, alors que la présence spécifique d'un ARNt, quelle que soit sa séquence, stimule cette activité. Cela suggère que l'hydrolyse de l'ATP est liée à la fixation ou la reconnaissance du substrat ARNt par le complexe KEOPS. Nous avons réalisé la même étude pour le complexe de KEOPS et Sua5 de *S.cerevisiae*, aboutissant à des résultats similaires.

Enfin, nous avons purifié les composants de la machinerie de synthèse de t<sup>6</sup>A chez *E.coli*, formé des protéines YrdC (orthologue de Sua5), YgjD (orthologue de Kae1, notée D), YjeE (notée E) et YeaZ (notée Z), et avons effectués des expériences de complémentation *in vitro* entre les systèmes de synthèse de t<sup>6</sup>A d'*E.coli*, de *S.cerevisiae* et de *P.abysyi*. Nous avons alors montré que les protéines YrdC/Sua5 d'une part, et les complexes KEOPS / DEZ d'autre part sont interchangeables, suggérant un mécanisme universel conservé pour ces différents systèmes enzymatiques.

**Ces résultats, associés aux données disponibles dans la littérature, nous ont permis de proposer un modèle pour la biosynthèse de t<sup>6</sup>A chez les Archées et les Eucaryotes (Figure ci contre), et l'ensemble de ces résultats et la discussion qui l'accompagne sont présentés dans l'article suivant.**



# *In vitro* biosynthesis of a universal t<sup>6</sup>A tRNA modification in Archaea and Eukarya

Ludovic Perrochia<sup>1</sup>, Estelle Crozat<sup>1</sup>, Arnaud Hecker<sup>1</sup>, Wenhua Zhang<sup>2</sup>, Joseph Bareille<sup>2</sup>, Bruno Collinet<sup>2,3</sup>, Herman van Tilbeurgh<sup>2</sup>, Patrick Forterre<sup>1,\*</sup> and Tamara Basta<sup>1,\*</sup>

<sup>1</sup>Institut de Génétique et Microbiologie, Université Paris-Sud, IFR115, UMR8621-CNRS, 91405 Orsay, France,

<sup>2</sup>Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, IFR115,

UMR8619-CNRS, 91405 Orsay, France and <sup>3</sup>UFR Sciences de la Vie, Université Pierre et Marie Curie UPMC, Sorbonne Universités, 75005 Paris, France

Received September 19, 2012; Revised November 9, 2012; Accepted November 11, 2012

## ABSTRACT

*N*<sup>6</sup>-threonylcarbamoyladenosine (t<sup>6</sup>A) is a modified nucleotide found in all transfer RNAs (tRNAs) decoding codons starting with adenosine. Its role is to facilitate codon–anticodon pairing and to prevent frameshifting during protein synthesis. Genetic studies demonstrated that two universal proteins, Kae1/YgjD and Sua5/YrdC, are necessary for t<sup>6</sup>A synthesis in *Saccharomyces cerevisiae* and *Escherichia coli*. In Archaea and Eukarya, Kae1 is part of a conserved protein complex named kinase, endopeptidase and other proteins of small size (KEOPS), together with three proteins that have no bacterial homologues. Here, we reconstituted for the first time an *in vitro* system for t<sup>6</sup>A modification in Archaea and Eukarya, using purified KEOPS and Sua5. We demonstrated binding of tRNAs to archaeal KEOPS and detected two distinct adenosine triphosphate (ATP)-dependent steps occurring in the course of the synthesis. Our data, together with recent reconstitution of an *in vitro* bacterial system, indicated that t<sup>6</sup>A cannot be catalysed by Sua5/YrdC and Kae1/YgjD alone but requires accessory proteins that are not universal. Remarkably, we observed interdomain complementation when bacterial, archaeal and eukaryotic proteins were combined *in vitro*, suggesting a conserved catalytic mechanism for the biosynthesis of t<sup>6</sup>A in nature. These findings shed light on the reaction mechanism of t<sup>6</sup>A synthesis and evolution of molecular

systems that promote translation fidelity in present-day cells.

## INTRODUCTION

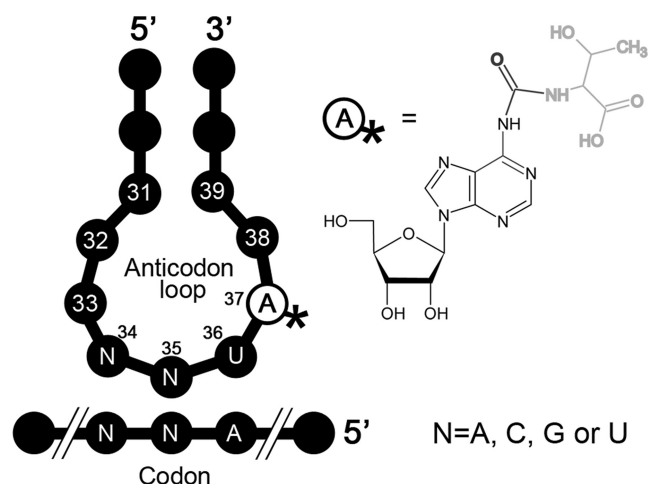
All naturally occurring cellular RNA species carry numerous modified nucleosides. Of 100 different modified nucleosides reported to date, >80 are present in mature transfer RNAs (tRNAs). Each tRNA molecule has some subset of these modifications, altering nucleoside identity at an average of 11 positions. Modified nucleosides favour correct folding and stability of tRNA, and some of them are important for the efficiency and fidelity of translation (1–5). Notably, some tRNA positions are almost always modified. This is the case for the nucleotide at the position 37, which is modified in >70% of all tRNAs (6). In a vast majority of cases, the modifications introduced at position 37 are the universal *N*<sup>1</sup>-methyl-guanosine (m<sup>1</sup>G) and *N*<sup>6</sup>-threonylcarbamoyladenosine (t<sup>6</sup>A), or their derivatives. t<sup>6</sup>A is found exclusively at the position 37, 3' adjacent to the third base of the anticodon, in all tRNAs reading ANN codons (where N corresponds to four canonical nucleotides A, C, G and T), tRNA<sup>Ile</sup>, tRNA<sup>Thr</sup>, tRNA<sup>Asn</sup>, tRNA<sup>Lys</sup>, tRNA<sup>Ser</sup> and tRNA<sup>Arg</sup> (Figure 1). The t<sup>6</sup>A modification facilitates correct anticodon–codon pairing by enhanced base stacking and by preventing intraloop base pairing of U<sub>33</sub> • A<sub>37</sub>, resulting in enhanced translation fidelity (7,8).

t<sup>6</sup>A modification has been discovered, and its chemical structure was elucidated >40 years ago (9). Initial studies of t<sup>6</sup>A biosynthesis performed *in vivo* or with *Escherichia coli* cellular extracts demonstrated that the reaction required adenosine triphosphate (ATP), carbonate,

\*To whom correspondence should be addressed. Tel: +33 1 69 15 74 89; Fax: +33 1 69 15 66 78; Email: patrick.forterre@igmors.u-psud.fr  
Correspondence may also be addressed to Tamara Basta. Tel: +33 1 69 15 46 28; Fax: +33 1 69 15 78 08; Email: tamara.basta@igmors.u-psud.fr  
Present address:

Arnaud Hecker, Université de Lorraine, UMR1136 INRA Université de Lorraine 'Interactions Arbres/Micro-organismes', IFR 110 EFABA, Faculté des Sciences BP 239, 54506 Vandoeuvre Cedex, France.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors



**Figure 1.**  $t^6A$  modification. On the left is a schematic representation of an anticodon loop of tRNA carrying NNU anticodon and an mRNA molecule carrying ANN codon. Adenosine at position 37 is indicated in white circle. On the right is the chemical structure of  $t^6A$ -modified nucleoside. Chemical groups originating from threonine or carbonate are indicated in grey and bold, respectively.

magnesium and threonine (10,11). However, the proteins involved in the catalysis of this multistep reaction were not identified until recently. The breakthrough came when de Crecy-Lagard *et al.* made the link between the universality of  $t^6A$  and a previous comparative genomic analysis that identified YrdC and YgjD and their archaeal and eukaryotic homologues (Sua5 and Kae1, respectively) as universal proteins of unknown function (12). In particular, they noticed that proteins HypF and NodU/CmcH, which catalyse chemical reactions similar to the ones expected in  $t^6A$  biosynthesis, are formed by the fusion of domains homologous to YrdC and YgjD. Using genetic experiments, they demonstrated that YrdC/Sua5 and YgjD/Kae1 are responsible for the formation of  $t^6A$  *in vivo* in *E. coli* and *Saccharomyces cerevisiae* (13,14).

Earlier this year, Deutsch *et al.* (15) successfully reconstituted the  $t^6A$  synthesis reaction *in vitro* by combining YrdC, YgjD and two additional *E. coli* proteins, YeaZ and YjeE, in the presence of ATP,  $Mg^{2+}$ , threonine and carbonate. Subsequently, the reconstitution of a *Bacillus subtilis* system using purified YwlC (YrdC), YdiB (YjeE), YdiC (YeaZ) and YdiE (YgjD) was reported (16). Notably, YeaZ and YjeE have no detectable orthologues in Archaea and Eukarya. YeaZ is a paralogue of YgjD (thus homologous but distantly related to Kae1), whereas YjeE is specific to Bacteria. This suggested either that Sua5 and Kae1 are sufficient to perform the reaction in these two domains of life or that the YeaZ and YjeE functions are performed by functional analogues evolutionarily unrelated to their bacterial counterparts. Obvious candidate proteins to fulfil this role are the partner proteins of Kae1 (the YgjD homologue) in the KEOPS (kinase, endopeptidase, and other proteins of small size) or EKC (endopeptidase-like and kinase associated with transcribed chromatin) complex. In addition to Kae1, the KEOPS/EKC complex (thereafter called KEOPS) contains three proteins conserved between

Archaea and Eukarya, the protein kinase Bud32, Cgi121 and Pcc1 (17,18). Genetic and transcriptomic studies in *S. cerevisiae* demonstrated that, in addition to Kae1, Bud32 and Pcc1 are required for the formation of  $t^6A$  modification (19,20). Recent genetic investigation of the KEOPS complex in halophilic archaeon *Haloferax volcanii* showed that Kae1, Bud32 and Cgi121 are essential, whereas Pcc1 knockout strain presented a decreased growth rate and reduced levels of  $t^6A$  modification (21). Furthermore, using a whole cell yeast extract for *in vitro*  $t^6A$  synthesis, Srinivasan *et al.* (20) showed that, as in the case of the bacterial system, the reaction requires Sua5, Kae1, threonine, bicarbonate and ATP.

We have previously characterized the archaeal homologues of Kae1 and Bud32, with the aim to get more insight into functions that were probably already present in the last universal common ancestor (LUCA; Kae1/YgjD) or in the last common ancestor of Archaea and Eukarya (Bud32). We have shown that Kae1 is not an *O*-sialoglycoprotein endopeptidase, as initially reported (22), but rather a metalloprotein that binds ATP and DNA *in vitro*, suggesting a role in some unknown universal function related to nucleic acid. In the course of our work, we have purified the archaeal KEOPS complex from the hyperthermophilic archaeon *Pyrococcus abyssi* (PaKEOPS, this work) and the *S. cerevisiae* KEOPS complex (ScKEOPS) (23) for structural and functional studies. Here, we report for the first time the *in vitro* formation of  $t^6A_{37}$ -modified tRNA catalysed by the archaeal and yeast complexes. We present in more detail the enzymatic properties of the more active PaKEOPS complex. The efficiency of the reconstituted *in vitro* system indicates that KEOPS and Sua5 are both necessary and sufficient for the reaction. Notably, PaKEOPS, but not Sua5, strongly binds tRNA *in vitro*, suggesting a role for KEOPS as a tRNA-binding platform in the  $t^6A$  modification reaction. Finally, we identified two distinct ATP-dependent catalytic steps associated with Sua5 and KEOPS. Our data, together with results obtained with the bacterial system (15), show that the universal proteins Sua5/YrdC and Kae1/YgjD cannot synthesize  $t^6A$  by themselves but require additional non-universal accessory proteins. Using interdomain complementation experiments, we demonstrated that the KEOPS complex and bacterial proteins YeaZ, YgjD and YjeE are functional analogues, suggesting mechanistic similarity of  $t^6A$  synthesis in the three domains of life. We discuss evolutionary implications raised by these observations and alternative models they suggested on the nature of LUCA.

## MATERIALS AND METHODS

### Cloning procedures

A polycistronic sequence containing the genes encoding the PaKEOPS subunits PAB1159 (Kae1), PAB1047 (Bud32), PAB1522 (Cgi121) and PAB3073 (Pcc1) was synthesized by GenScript and cloned into pET28(a) plasmid (Novagen). PAB3073 (Pcc1) was fused to a hexa-histidine tag at its 3'-end. The gene PAB1159



(KaeI) was cloned into the pET28(a) as previously described (24).

The gene PAB1302 encoding PaSua5 was amplified by polymerase chain reaction (PCR) using total genomic DNA from the *P. abyssi* GE5 strain as template and cloned into pET26b (Novagen). The ScSua5 gene was amplified by PCR using genomic DNA from *S. cerevisiae* S288C strain (ATCC). A hexa-histidine tag was added at the 3'-end of the gene, which was subsequently cloned into pET21d vector (Novagen).

*E. coli* genes *ygjD* and *yeaZ* were cloned in pET9a with a histidine tag at their 3'-end; *yrdC* was cloned in pET26b with a histidine tag at its 3'-end. *yjeE* gene was synthesized (GenScript) and cloned in pET28a with a Strep-tag II on its 3'-end.

Media were supplemented with 50 µg/ml kanamycin, 34 µg/ml chloramphenicol or 100 µg/ml ampicillin when needed.

### Recombinant protein expression and purification

Recombinant proteins were expressed in Rosetta2 (DE3) pLysS *E. coli* strains (Novagen). Protein overexpression was induced in Overnight Express Instant TB medium (Novagen) supplemented with 10% glycerol.

Cells were collected by centrifugation, resuspended in lysis buffer (Lbf) and sonicated on ice. After centrifugation and in the case of archaeal protein purification, the supernatant was heated at 65°C for 15 min to precipitate bacterial proteins. His-tagged proteins from the soluble fraction were purified by gravity-flow chromatography on an Ni-NTA resin column (Qiagen) according to the manufacturer's recommendations. Proteins carrying a strep-tag II were purified on a Strep-tactin Sepharose resin (IBA) following manufacturer advice. Fractions of interest were pooled, concentrated and injected on a Superdex<sup>TM</sup> 200 column (GE Healthcare) for KEOPS or Superdex<sup>TM</sup> 75 column for Sua5 and bacterial proteins. Fractions containing pure proteins were concentrated and stored at -20°C in Lbf without β-mercaptoethanol and 10% glycerol.

### tRNA expression plasmid construction

*P. abyssi* (Pa) and *E. coli* (Ec) tRNA gene sequences PaVal (GAC), PaLys (UUU), PaIle (GAU), PaAsn (GUU), PaMet (CAU), EcVal (GAC), EcLys (UUU), EcIle (GAU) and EcAsn (GUU) were retrieved from the Genomic tRNA Database (<http://gtrnadb.ucsc.edu/>) (25). Two complementary oligonucleotides corresponding to the forward and reverse sequence of tRNA genes were synthesized (IDT), annealed and cloned in a pBlue-Script vector yielding pB-tRNA, to allow their expression under the control of T5 promoter/*lac* operator sequence. All sequences are available on request.

### tRNA production and purification

XL1 Blue *E. coli* cells (Agilent Technologies) were transformed with pB-tRNA, and tRNAs were overexpressed in Overnight Express Instant TB medium (Novagen). After centrifugation, total RNAs were extracted with Trizol (Sigma) using the manufacturer's procedure.

Run-off transcription of tRNA genes was performed according to Mougin *et al.* (26): Pa\_tRNA<sup>Lys</sup> and Ec\_tRNA<sup>Asn</sup> genes were cloned in pB-tRNA and amplified by PCR using primers containing the sequence of the T7 promoter upstream of the gene to be transcribed. The transcription reaction was carried out in transcription buffer supplemented with 200 nmol of each rNTP, 40 units of RNasin (Thermo Scientific), 300 ng of PCR product and 0.2 µM of his-T7 RNA polymerase expressed and purified according to Arnaud *et al.* (27). Reaction occurred at 37°C for 4 h. Template DNA was then digested with DNase I (Thermo Scientific) for 30 min at 37°C.

For purification, RNAs were loaded on an 8 M urea, 12% polyacrylamide gel, and the dominant band corresponding to overexpressed tRNA was cut out. tRNA was eluted from gel in elution buffer, precipitated and resuspended in water. To favour correct folding of tRNAs, the solution was heated at 70°C, cooled slowly to room temperature and stored at -80°C.

### In vitro assay for the synthesis of t<sup>6</sup>A-modified tRNA

The reaction was performed in a final volume of 50 µl containing archaeal (1 µM or 5 µM for complementation experiments), yeast (5 µM) or bacterial (5 µM) proteins, tRNA (2 µM) and 18.2 µM <sup>14</sup>C-L-threonine (0.05 µCi, 55 Ci/mol, Hartmann Analytic) in reaction buffer. After 30-min incubation at 50°C (or other incubation times when indicated), macromolecules were precipitated by the addition of 1 ml of trichloroacetic acid 15% (TCA), and incubated at 4°C for 1 h. Precipitated material was applied on pre-wet glass microfibre filters GF/F (Whatmann) using vacuum apparatus (Millipore). Filters were washed with 2 ml of 5% TCA and 3 ml of 95% EtOH and dried. Radioactivity was recorded as average counts per minute (CPM) for 2 min, with a Packard liquid scintillation analyser.

The amount of incorporated threonine was calculated from a standard curve obtained with different concentrations of <sup>14</sup>C-L-threonine spotted on filters (1 pmol = 102.8 CPM).

Yield was calculated as [CPM (assay) - CPM (background)]/102.8]/(pmol tRNA) × 100.

### 2D thin layer chromatography for detection of t<sup>6</sup>A-modified nucleosides

tRNAs obtained after *in vitro* t<sup>6</sup>A reaction were extracted with phenol/chloroform and ethanol precipitated. The tRNA pellet was dissolved in 10 µl of 100 mM ammonium acetate supplemented with 1 µg of P<sub>1</sub> nuclease from *Penicillium citrinum* (Sigma) and incubated overnight at 37°C. Digested tRNAs (2 µl) were mixed with 20 µg of cold 5'-mononucleosides pA, pU, pG and pC and spotted on a CEL-300 cellulose plate (Merck). 5'-mononucleosides were separated using chromatographic solvents A, B or C as previously described (28). First dimension chromatography was performed in solvent A, and the second in solvent B or C. The positions of the four major mononucleosides, pA, pG, pU and pC, were revealed by UV shadowing. The position of t<sup>6</sup>A nucleoside

carrying radiolabelled threonine was revealed by phosphorimaging and compared with that of reference maps obtained under identical experimental conditions.

#### ATP hydrolysis assay

ATP hydrolysis reactions were carried out in a final volume of 25  $\mu$ l containing archaeal or yeast proteins (2  $\mu$ M, if not otherwise specified) in reaction buffer, supplemented with 100  $\mu$ M of cold ATP and 1  $\mu$ Ci of  $\alpha$ - $^{32}$ P-ATP (PerkinElmer). When indicated, tRNA, single- or double-strand random sequence RNA (1  $\mu$ M) and other amino acids (50  $\mu$ M) were added to the reaction.

Reaction was performed at 50°C during 30 min, stopped on ice, and 1  $\mu$ l of the mix was spotted on a PEI cellulose plate (Merck), pre-run in distilled water. Radioactive nucleotides were separated by thin layer chromatography (TLC) using 0.5 M  $\text{KH}_2\text{PO}_4$ , pH 3.5, as the solvent. Plates were dried and revealed by phosphorimaging.

#### Electrophoretic mobility shift assay

The binding reactions were carried out in a final volume of 20  $\mu$ l containing 0.05 pmol (10 nM) of RNA probes and purified PaKEOPS or PaSua5 (0.5 nM to 50  $\mu$ M) in binding buffer. Incubations were performed for 15 min at 4°C, the products were loaded onto a 10% non-denaturing acrylamide gel and electrophoresis was performed in 1  $\times$  TGE buffer at 4°C for 4 h at 7.5 V/cm. The RNA–protein complexes were visualized by autoradiography and phosphorimaging.

## RESULTS

### *P. abyssi* KEOPS binds tightly to tRNAs

The whole PaKEOPS complex (Kae1, Bud32, Cgi121 and Pcc1) was purified from an *E. coli* overproducing strain by exploiting the polyhistidine tag of the Pcc1 subunit. The complex eluted as a single peak with an apparent molecular mass of  $\sim$ 200 kDa after gel filtration (Supplementary Figure S1A). The theoretical mass of PaKEOPS was 90.5 kDa; therefore, the apparent molecular mass of this peak could correspond to a dimer, in agreement with the 3D structural model proposed previously (29). Interestingly, PaKEOPS co-purified with nucleic acids that were resistant to DNaseI digestion but sensitive to RNaseA (not shown) and displayed a similar electrophoretic migration pattern as tRNAs isolated from *E. coli* expression cells (Supplementary Figure S1B). These observations suggested that purified recombinant PaKEOPS was bound to *E. coli* tRNAs. In contrast, recombinant PaSua5 did not co-purify with nucleic acids.

Electrophoretic mobility shift assays of PaSua5 and PaKEOPS in the presence of substrate, non-substrate tRNA and random sequence 65mer RNA showed binding of all RNAs only by PaKEOPS, whereas PaSua5 did not bind tRNA even at a high protein to tRNA molar ratio (Supplementary Figure S2). This suggested that the discrimination between substrate and non-substrate tRNAs occurs after the binding of tRNA to the complex and that specific tRNA-like structural and

sequence features are not required for the binding of RNA by KEOPS. Taken together, these observations suggested that PaKEOPS might be mainly responsible for the binding of tRNA substrates during t<sup>6</sup>A synthesis.

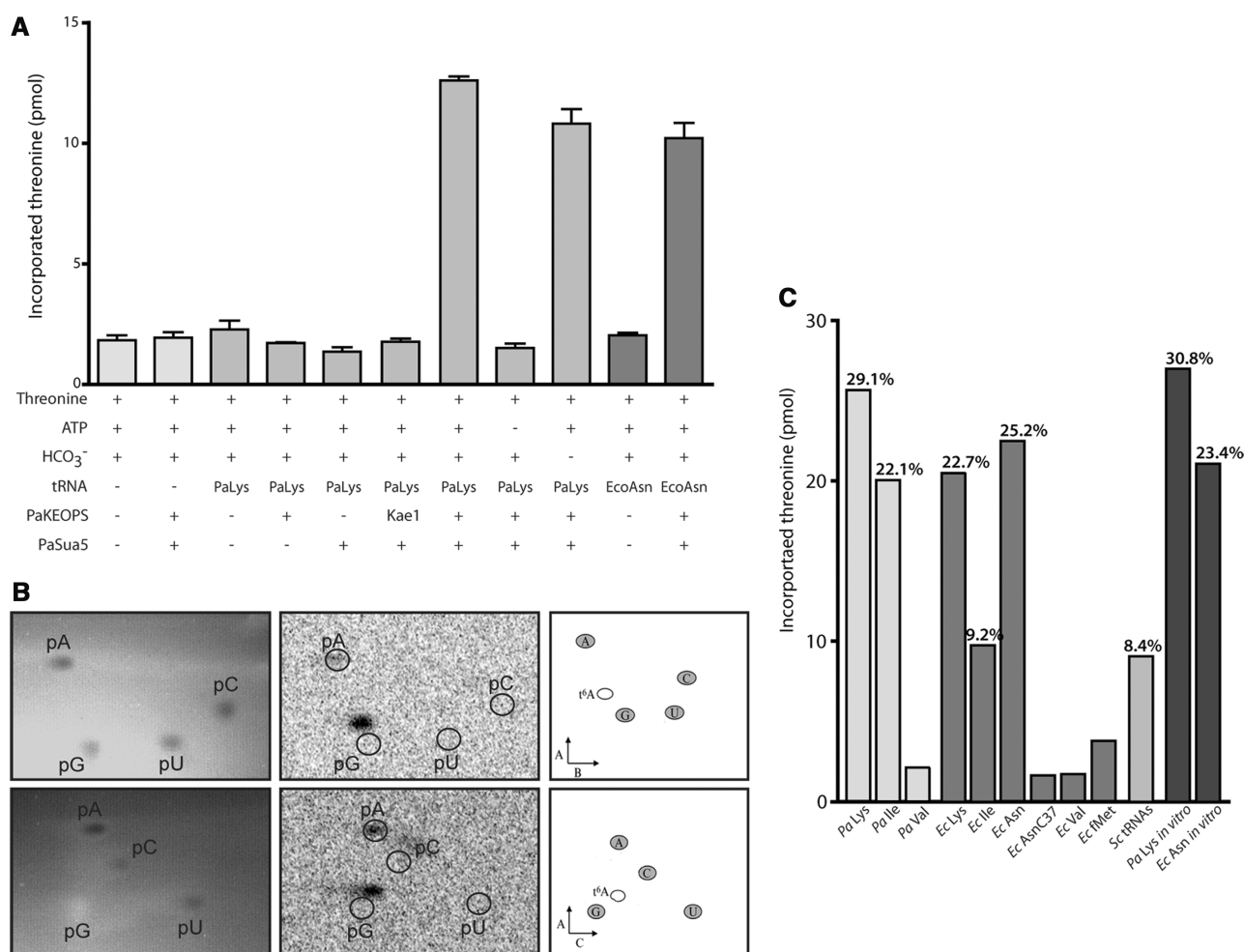
### PaSua5 and PaKEOPS together catalyse the biosynthesis of t<sup>6</sup>A

To determine whether PaSua5 and PaKEOPS were able to catalyse the biosynthesis of t<sup>6</sup>A, we needed an appropriate tRNA substrate lacking this modification. We initially anticipated that additional modifications might also be important for the formation of t<sup>6</sup>A because a previous *in vivo* study using *Xenopus* oocytes showed that t<sup>6</sup>A and several nucleotide modifications (D47, m5C48, m1A58) were formed almost simultaneously (30). To obtain tRNA substrates that most closely satisfied these criteria, we designed a recombinant strategy for tRNA production *in vivo* using *E. coli* cells that overexpressed various *P. abyssi* (Pa\_tRNA) or *E. coli* (Ec\_tRNA) tRNA genes under the control of an inducible promoter. Under these conditions, it is known that the resulting overexpressed tRNA is often hypomodified, thus generating a collection of so-called modivariants (RNA molecules that differ only by their modified nucleoside content) (31,32). Using this system, we typically obtained a major band on denaturing acrylamide gel corresponding to overexpressed, possibly undermodified tRNAs, species (Supplementary Figure S3A).

We incubated PaSua5 and PaKEOPS in presence of ATP, Pa\_tRNA<sup>Lys</sup>, carbonate and  $^{14}$ C-threonine, and measured the incorporation of  $^{14}$ C-threonine into TCA-precipitated material. We observed a significant increase of radioactivity when both PaSua5 and PaKEOPS were present in the reaction mixture (Figure 2A). In contrast, no incorporation was observed with PaSua5 or PaKEOPS alone, indicating that both are necessary for the reaction to take place. Notably, PaSua5 and PaKae1 together do not catalyse the incorporation of threonine, indicating that PaKae1 cannot substitute for PaKEOPS. Threonine incorporation also required the addition of ATP and tRNA but, surprisingly, not carbonate. We hypothesized that carbonate was still present in the reaction mixtures despite boiling and extensive degassing of reaction solutions. For comparison, we performed the complete t<sup>6</sup>A assay using *E. coli* proteins (as explained later). As shown in Supplementary Figure S7C, added carbonate had a stimulating effect on the *E. coli* system, indicating that at least some of atmospheric carbonate was removed from reaction solutions. Thus, it appears that archaeal and also eukaryal systems (as explained later) function at lower carbonate concentrations as compared with bacterial systems.

The optimal temperature for threonine incorporation was 50°C with 65% of activity still retained at 73°C (Supplementary Figure S3C). Because the optimal growth temperature of *P. abyssi* GE5 is 96°C, it is likely that the reaction is limited by thermal denaturation of tRNA in non-physiological *in vitro* conditions.

Importantly, the threonine incorporation test was negative for non-ANN decoding tRNA<sup>Val</sup> from *E. coli*



**Figure 2.** *In vitro* reconstitution of t<sup>6</sup>A-modified tRNAs. (A) Incorporation of radiolabelled threonine in tRNAs: reaction mixtures composed of the reactants indicated under the graph were precipitated with TCA, and the amount of incorporated <sup>14</sup>C-threonine was measured. The tRNA species from *P. abyssi* (Pa) and *E. coli* (Ec) (purified from *in vivo* overexpression) are indicated. In one case, Kae1 was used instead of the whole KEOPS complex. HCO<sub>3</sub><sup>-</sup> was added (+) or not (-) to the reaction. Error bars indicate standard deviation from the mean obtained for three independent experiments. (B) Detection of t<sup>6</sup>A-modified nucleotide using 2D TLC. Upper line shows chromatographic mobility in solvent system A + B, lower line in solvent system A + C. Left panels: abundant, non-modified nucleotides were detected by ultraviolet (UV) shadowing. Centre panel: <sup>14</sup>C-threonine incorporated in modified nucleotides was detected by phosphorimaging, and the positions of unmodified nucleotides were superimposed by aligning with the UV image (white circles). Right panel: reference maps [from (28)] showing the theoretical position of t<sup>6</sup>A nucleoside in both solvent systems. (C) Incorporation of radiolabelled threonine using different tRNA species as substrates. Legend is the same as in (A); tRNAs were overexpressed *in vivo*, or produced *in vitro* when stated. fMet, bacterial initiator *N*-formyl-methionine; AsnC37 is Asn tRNA where A37 was mutated into a cytosine; Sc tRNAs are tRNAs purified from a  $\Delta$ sua5 *S. cerevisiae* strain. The percentage indicated on top of each bar is the yield of t<sup>6</sup>A-modified tRNA obtained for each reaction.

and *P. abyssi*, thus confirming that only ANN-decoding tRNA molecules are valid substrates. Moreover, Ec\_tRNA<sup>Asn</sup> failed to act as a substrate when adenosine at position 37 was replaced by cytosine, showing that t<sup>6</sup>A is specifically added on the adenosine 37.

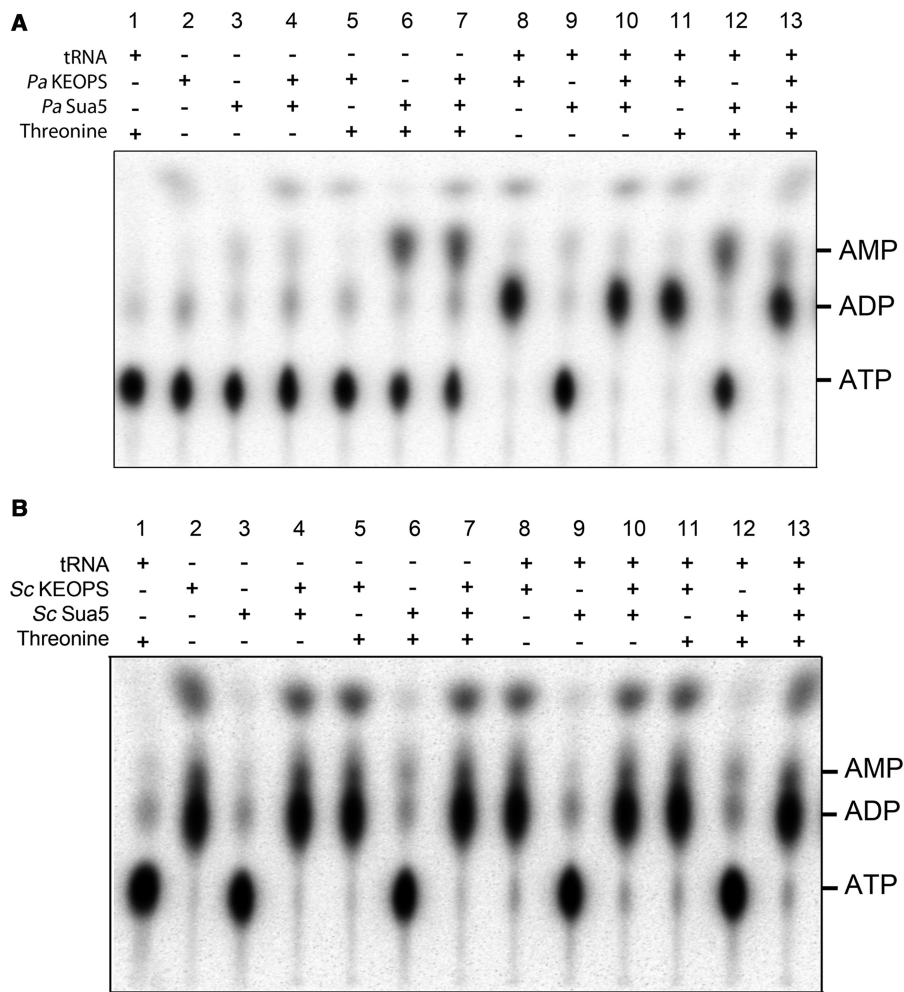
The incorporation of threonine into tRNAs was confirmed by purification of tRNAs after t<sup>6</sup>A reaction. As expected, only tRNAs from the reaction containing PaSua5 and PaKEOPS were radioactive (Supplementary Figure S3B).

Finally, we performed 2D TLC to demonstrate that the incorporation of threonine into tRNA corresponded to t<sup>6</sup>A formation. For this, tRNAs were isolated from the reaction mixture, hydrolysed to individual nucleotides and separated by 2D TLC. The position of the radioactive

spot on the plate was reproducible and consistent with the position of the t<sup>6</sup>A-modified nucleotide according to previously published reference maps (33) (Figure 2B).

Next, it was important to verify that the *in vitro* assay we established can generate a significant amount of t<sup>6</sup>A-modified tRNA. Standard assay conditions yielded 25–32% of modified Pa\_tRNA<sup>Lys</sup> (i.e. mole t<sup>6</sup>A<sub>37</sub> per mole of tRNA present in the incubation mixture). This result indicated that the *in vitro* reaction could still be optimized or that the enzymatic machinery may work better *in vivo* under the physiological conditions of the cell. To optimize the reaction, we added manganese divalent ions, which were reported to be required for optimal activity of Bud32 (34). Indeed, manganese had a stimulating effect because we could increase the yield of





**Figure 3.** ATPase activity of Sua5 and KEOPS. (A) *P. abyssi* proteins and (B) *S. cerevisiae* proteins. Produced radiolabelled ( $\alpha$ - $^{32}$ P) ADP and AMP were separated by TLC and visualized by phosphorimaging. Presence (+) or absence (–) of different reactants in reaction mixtures is indicated. Pa\_tRNA<sup>Lys</sup> produced in overexpressing *E. coli* strain was used as a substrate.

the reaction to 50% for archaeal system and to 38% for the *S. cerevisiae* system (as explained later) (Supplementary Figure S5).

However, the yield reported is the lowest estimate for our assay conditions because, as noted previously, it is possible that a subpopulation of the *in vivo* produced Pa\_tRNAs<sup>Lys</sup> already carried the t<sup>6</sup>A modification. The yield could not be improved by adding a fresh aliquot of PaSua5 and PaKEOPS or ATP into the reaction mixture (not shown), which indicated that the amount of substrate tRNA was the limiting factor.

To further characterize the range of substrate tRNAs used by PaKEOPS and PaSua5, we also tested overexpressed *E. coli* tRNAs, total tRNAs isolated from a *S. cerevisiae* Sua5 deletion mutant (19) and *in vitro* transcribed tRNAs. All of these acted as equivalent substrates (Figure 2C), demonstrating that (i) universally conserved structural and/or sequence features of tRNAs decoding ANN codons are recognized by PaKEOPS and PaSua5; and (ii) presence of modified nucleotides on ANN-decoding tRNA is not required for the t<sup>6</sup>A<sub>37</sub> modification to take place.

**Two distinct ATP-dependent reactions occur in the course of t<sup>6</sup>A synthesis by PaSua5 and PaKEOPS**

To investigate the ATP-dependent activation steps undertaken by PaKEOPS and PaSua5, we assayed the ATPase activity using  $\alpha$ - $^{32}$ P-ATP and TLC to visualize the resulting products (Figure 3A).

When all the proteins and substrates were put together in a reaction that leads to t<sup>6</sup>A synthesis and under saturating concentration of ATP, we observed that both AMP and ADP were produced. The production of AMP occurred when PaSua5 was present in the reaction mixture alone. This activity was significantly increased when threonine and carbonate were added but did not require tRNA (Figure 3A and Supplementary Figure S7A and S7B). In contrast, PaSua5 activity was not stimulated by glycine, phenylalanine and homoserine, an isomer of threonine (Supplementary Figure S4B).

In contrast to PaSua5, incubation of PaKEOPS alone with ATP generated ADP (Figure 3A). This basal activity is detected in the absence of tRNA; however, the addition of substrate tRNA or non-substrate Ec\_tRNA<sup>Asn</sup> (C37) and Pa\_tRNA<sup>Val</sup> significantly increased this activity

(Figure 3A and Supplementary Figure S4A). Notably, such activation was not observed when tRNA was replaced by random sequence 65mer DNA or RNA, although this RNA was bound by KEOPS with an affinity similar to that observed for tRNA (see Supplementary Figures S2C and S4A).

The KEOPS-dependent ATP hydrolysis into ADP seems to be required for t<sup>6</sup>A synthesis because the t<sup>6</sup>A reaction was completely inhibited when ATP was replaced by AMP-PNP, an ATP analogue that can be hydrolysed into AMP (data not shown).

Taken together, our data indicate that two distinct ATP-dependent reactions, one generating AMP and another generating ADP, are catalysed by PaSua5 and PaKEOPS, respectively, in the course of t<sup>6</sup>A synthesis.

#### *In vitro* synthesis of t<sup>6</sup>A by Sua5 and KEOPS from *S. cerevisiae*

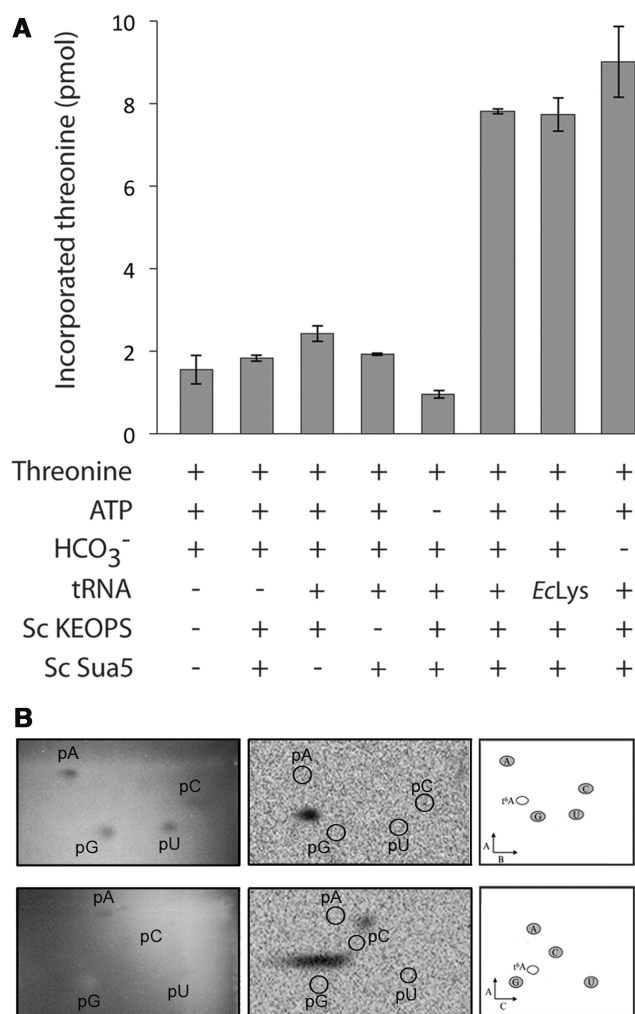
The *S. cerevisiae* KEOPS (ScKEOPS) complex was heterologously expressed and purified as previously described (23). In contrast to PaKEOPS, ScKEOPS was not associated with tRNAs when purified from *E. coli* cells (not shown). Consistent with observations made for PaSua5, the purified recombinant yeast homologue ScSua5 did not co-purify with tRNA either.

To test whether ScSua5 and ScKEOPS could also catalyse the biosynthesis of the t<sup>6</sup>A modification, we used the same experimental approach as for *P. abyssi* proteins except that the assay was performed at 30°C. As for *P. abyssi* proteins, we observed the incorporation of <sup>14</sup>C-threonine into TCA-precipitated material, and 2D TLC analysis confirmed that t<sup>6</sup>A was formed (Figure 4B). This reaction was ATP and tRNA dependent and required the presence of ScSua5 and ScKEOPS. Similar as for the *P. abyssi* system, the omission of carbonate did not have an effect on the amount of incorporated threonine (Supplementary Figure S7C). The amount of threonine incorporation was similar for endogenous tRNAs and *E. coli* tRNAs, indicating that, like *P. abyssi* proteins, the *S. cerevisiae* proteins recognize features common to all ANN-decoding tRNA molecules.

We next examined the ATPase activity of ScSua5 and ScKEOPS using TLC (Figure 3B). We observed weak ATPase activity of ScSua5 of which the product was AMP. This activity was threonine dependent and carbonate dependent (Supplementary Figure S7A and S7B) but did not require tRNA, and addition of tRNA did not have a stimulating effect. Thus, the ScSua5 behaved in the same manner as the PaSua5. In contrast, the behaviour of ScKEOPS was different from that of PaKEOPS because we observed the production of ADP and AMP in a tRNA-independent manner. Note, however, that we cannot exclude a contaminating ATPase activity from the *E. coli* expression strain.

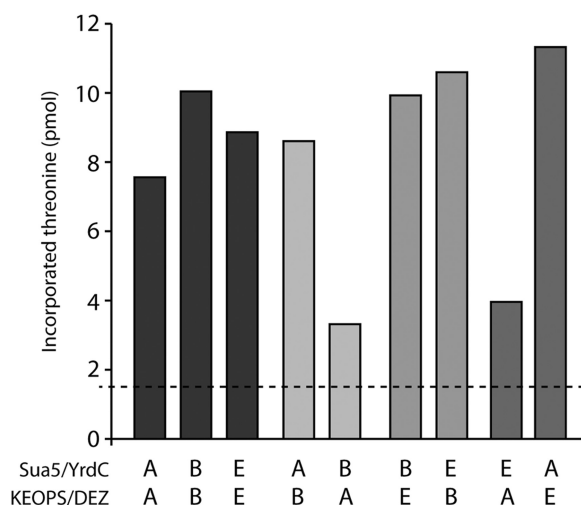
#### Interdomain exchange of t<sup>6</sup>A machinery components

The similar behaviour of archaeal, yeast and bacterial t<sup>6</sup>A machineries suggested that the function of KEOPS/YgjD-YjeE-YeaZ (hereafter called DEZ) and Sua5/YrdC has been conserved throughout evolution. To test this



**Figure 4.** *In vitro* synthesis of t<sup>6</sup>A modification using *S. cerevisiae* KEOPS and Sua5 proteins. (A) Incorporation of radiolabelled threonine in tRNAs: the amount of incorporated <sup>14</sup>C-threonine was measured as in Figure 2A. Total tRNAs were purified from a  $\Delta$ Sua5 strain; for comparison, *in vitro* transcribed *E. coli* tRNA<sup>Lys</sup> was used when indicated. Error bars indicate standard deviation from the mean obtained from three independent experiments. (B) Detection of t<sup>6</sup>A-modified nucleotide using 2D TLC. The chromatography was done as described in Figure 2B.

hypothesis, we purified bacterial proteins and set up a functional *in vitro* system to perform complementation assays. We combined Sua5/YrdC proteins and KEOPS/DEZ complexes. TLC analysis confirmed that the mixed systems produce t<sup>6</sup>A-modified tRNA (Supplementary Figure S6). Remarkably, despite the non-optimal temperatures used in the assays, the amount of t<sup>6</sup>A produced by heterologous combinations was comparable with the amount produced by the homologous systems at the respective reaction temperatures (43°C for Pa and Ec proteins, and 35°C for Sc proteins) (Figure 5). Lower values were obtained in complementation with PaKEOPS, owing to suboptimal temperatures because at 43°C and 35°C, PaKEOPS retains 40% and 25% of its activity (Supplementary Figure S3C). Nevertheless, those values were reproducible and are still fairly above the background signal. These results indicate that Sua5/YrdC



**Figure 5.** Functional complementations assays between bacterial, archaeal and eukaryotic proteins. Complementation assays were tested using PaKEOPS, ScKEOPS or EcDEZ combined with PaSua5, ScSua5 or EcYgjD. Incorporation of radiolabelled threonine was measured as in Figure 2A. Activities of archaeal (A), bacterial (B) and eukaryotic (E) complexes were measured at 43°C (A, B and A/B) and 35°C (E, B/E and A/E). The dotted line indicates the background level, obtained without ATP.

and KEOPS/DEZ components can be exchanged, suggesting that t<sup>6</sup>A-synthesizing machineries from the three domains of life share a common catalytic mechanism.

## DISCUSSION

### Archaeal and eukaryal Sua5 and KEOPS together catalyse the synthesis of t<sup>6</sup>A-modified tRNA

In the present study, we established an *in vitro* system for enzymatic synthesis of a universal and essential t<sup>6</sup>A tRNA modification by archaeal and eukaryotic KEOPS and Sua5 proteins. This work thus provides for the first time direct evidence for the biological function of KEOPS and Sua5 in archaeal and eukaryotic cells and confirms the results of previous genetic and transcriptome studies in *S. cerevisiae* and *H. volcanii* (13,14,19,20,21). In the light of these findings, it is possible that the pleiotropic phenotypes described for yeast KEOPS and Sua5 deletion mutants (17,18,35) are indirect effects owing to the translational defects occurring in cells lacking t<sup>6</sup>A modification. This is well illustrated by a recent study that showed that the diverse cellular dysfunctions linked to the development of type 2 diabetes in mice can be attributed to abnormal protein synthesis caused by the lack of 2-methylthio-N<sup>6</sup>-threonylcarbamoyladenosine (ms<sup>2</sup>t<sup>6</sup>A)-modified tRNA<sup>Lys</sup> (36).

Purified proteins and a reconstituted *in vitro* system allowed us to get the first insights onto the structural and mechanistic features of the t<sup>6</sup>A modification process. We made two important observations: firstly, PaKEOPS, but not PaSua5, binds tRNA tightly; and secondly, the reaction requires the hydrolysis of ATP into AMP by Sua5 and the hydrolysis of ATP into ADP by KEOPS.

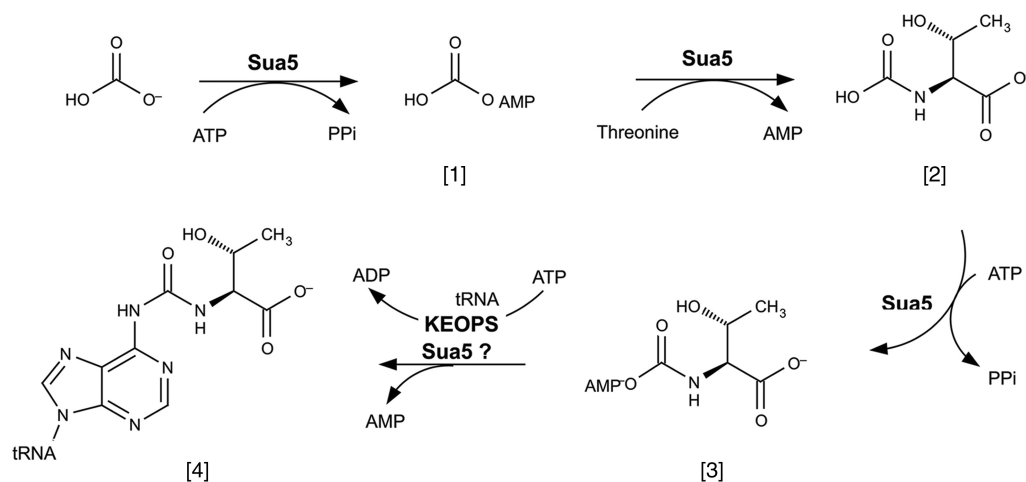
### The KEOPS complex is a tRNA binding platform

PaKEOPS binds tRNA of *E. coli* *in vivo* and *in vitro*, and this binding is strong because it is preserved throughout the purification process. It therefore seems plausible, suggesting that the KEOPS complex is responsible for tRNA binding in the course of the t<sup>6</sup>A synthesis. Interestingly, PaKEOPS binds tRNA in a positive cooperative manner (Supplementary Figure S2). It is unlikely that PaKEOPS binds more than one molecule of tRNA simultaneously because of the large size of tRNA. An alternative explanation is that one tRNA molecule is bound by each subunit of a PaKEOPS homodimer. We have previously shown that Kael from *P. abyssi* binds DNA *in vitro* (24). This suggested that Kael could have a general affinity for nucleic acids. Indeed, we could show using electrophoretic mobility shift assays that purified *P. abyssi* Kael can bind ssRNA and tRNA, albeit with lower affinity as compared with KEOPS (unpublished data). This suggests that tRNA interacts with Kael in the KEOPS/tRNA complex. However, because purified Kael is not bound to *E. coli* tRNA, it is likely that other proteins of the complex are also involved in tRNA binding. Interestingly, the archaeal/eukaryal KEOPS and bacterial t<sup>6</sup>A machinery can use as substrate tRNA from the three domains of life. This again suggests that Kael is directly involved in tRNA binding because it is the only KEOPS protein with a bacterial homologue (YgjD). This also indicates that universal structural and/or sequence features of tRNA are recognized by all systems, suggesting a similar mode of interaction between tRNA and the proteins.

### The catalytic mechanism for t<sup>6</sup>A synthesis seems to be conserved in the three domains of life

It was suggested that t<sup>6</sup>A synthesis, which is a complex modification, required at least two ATP-dependent activation steps (37). In agreement with this proposal, Deutsch *et al.* reported that both AMP and ADP were produced when all four proteins of the bacterial system were mixed (15). They observed that YrdC alone hydrolysed ATP into AMP in a threonine-dependent reaction, and that YgjD, YjeE and YeaZ (DEZ) generated ADP. These results were confirmed by a recent study that reconstituted *in vitro* the t<sup>6</sup>A biosynthesis by *B. subtilis* proteins YwLC (YrdC), YdiB (YjeE), YdiC (YeaZ) and YdiE (YgjD) (16). One common point between the bacterial system and our archaeal/eukaryal systems is the production of both AMP and ADP in the reaction that generates t<sup>6</sup>A-modified tRNA. The conservation of this feature across the three domains of life corroborates the hypothesis of at least two distinct ATP-dependent steps in the course of t<sup>6</sup>A synthesis, suggesting a conserved catalytic mechanism for the biosynthesis of t<sup>6</sup>A in nature, despite the fact that some of the proteins are unrelated. The second common point of all three systems is the carbonate and threonine dependence of AMP generation by YrdC/Sua5. This observation is in line with the universal character of these proteins and indicates that they perform the same catalytic role in the three domains of life.





**Figure 6.** Proposed reaction scheme for  $t^6A$  synthesis. Different catalytic steps are catalysed by proteins indicated above arrows. Activation of bicarbonate to carboxy-AMP [1]; condensation of carboxy-AMP with threonine to form carbamoylthreonine [2]; activation of carbamoylthreonine to threonylcarbamoyladenylate [3]; transfer of threonylcarbamoyladenylate to Kael/YgjD active site where second condensation reaction occurs to form  $t^6A$ -modified tRNA [4]. The participation of Sua5 in the last catalytic step is currently unclear. Further details are in the text.

The three systems diverge significantly in terms of complexes involved in the ADP-generating reaction step. In Bacteria, this reaction is catalysed by DEZ, whereas it is catalysed by KEOPS in Archaea and Eukarya. Kael/YgjD is the only common protein between these different systems, suggesting that it might be responsible for this activity.

Finally, using *in vitro* complementation experiments, we have demonstrated that Sua5 can replace YrdC and *vice versa* and that KEOPS is interchangeable with DEZ. These data corroborate the notion of a common catalytic mechanism for  $t^6A$  synthesis catalysed by universal proteins Sua5/YrdC and Kael/YgjD.

#### A putative mechanism for $t^6A$ synthesis

The threonine-dependent hydrolysis of ATP into AMP by Sua5 and YrdC suggests the formation of an adenylate reaction intermediate, such as threonylcarbamoyladenylate. The fact that we do not directly observe threonylcarbamoyladenylate is probably because of its unstable nature under acidic conditions used for TLC analysis. This intermediate was isolated using HPLC under neutral pH conditions with *B. subtilis* YwlC (YrdC), which catalysed the conversion of L-threonine, bicarbonate/ $CO_2$  and ATP to give L-threonylcarbamoyl-AMP intermediate and pyrophosphate as products (16). Moreover, L-threonylcarbamoyl-AMP was found in the 3D crystal structure of Sua5 from *Sulfolobus tokodaii*, and a recent study of TobZ protein also suggested a putative model for  $t^6A$  synthesis involving a threonylcarbamoyladenylate intermediate (38,39). This protein, which synthesizes the antibiotic nebramycin, consists of two domains homologous to YrdC and Kael fused into a single polypeptide. In the proposed model for the reaction catalysed by TobZ, the YrdC-like domain catalyses the ATP-dependent formation of the carbamoyladenylate intermediate, followed by the transfer of the carbamoyl moiety to tobramycin to form nebramycin with

concomitant release of AMP. This final step is performed by the Kael-like domain of TobZ, which positions the adenylate intermediate and the sugar substrate in the correct orientation. By analogy, Parthier *et al.* (39) suggested that Sua5 catalyses the formation of threonylcarbamoyladenylate, which is then accommodated in the active site of Kael, together with the anticodon loop of tRNA, to form  $t^6A$ -modified tRNA. Our results are compatible with this model. However, Parthier *et al.* further suggested that N-carbamoylthreonine, the precursor for the synthesis of the threonylcarbamoyladenylate intermediate, is generated from bicarbonate and threonine by an unknown enzymatic activity. Our data prove that it is not the case because Sua5 and KEOPS are sufficient to synthesize  $t^6A$ . Moreover, the observation that ATPase activity of PaSua5 [and of YrdC and YwlC in bacterial systems (15,16)] is carbonate and threonine dependent strongly suggests the involvement of Sua5/YrdC in the early steps of the catalysis, leading to the formation of threonylcarbamoyladenylate. In addition, our data imply that KEOPS is involved in the last step of the catalysis, which requires the presence of tRNA. This is supported by the finding that the reaction of tRNA with L-threonylcarbamoyl-AMP can be catalysed by the *B. subtilis* YdiB (YjeE), YdiC (YeaZ) and YdiE (YgjD) complex in the absence of YwlC (YrdC) (16). Based on these data and previous studies, we propose a novel mechanism for the synthesis of  $t^6A$  (Figure 6). In the first step, bicarbonate is activated by Sua5/YrdC to carboxy-AMP [1], which can then condense with the threonine to form carbamoylthreonine [2]. Sua5/YrdC uses a second ATP to activate this product and generates threonylcarbamoyladenylate [3]. Finally, this intermediate is transferred to Kael/YgjD active site where second condensation reaction occurs to form  $t^6A$ -modified tRNA. This proposal is compatible with the data reported by Lauhon (16), but it is in disagreement with the work of El Yacoubi *et al.* (14), who proposed the activation of

bicarbonate to carboxyphosphate catalysed by KaeI. We do observe generation of ADP by KEOPS, and this activity is mandatory. However, KEOPS activity is sensible to neither carbonate nor threonine, suggesting strongly that the complex, and thus KaeI, is not involved in early steps of catalysis. In favour of our proposal, YrdC-homologous domain of TobZ protein was shown to catalyse formation of carbamoyl-AMP, a compound that is structurally reminiscent of carboxy-AMP. Our model is compatible with the data obtained for the bacterial t<sup>6</sup>A systems, and carboxy-AMP was included in the proposed mechanism scheme as an alternative to the formation of carboxyphosphate (15). The present data do not allow assigning convincingly the ADP-generating ATPase activity of KEOPS/DEZ to a particular process. However, in archaeal system, we observed that the active site responsible for ATP hydrolysis is not in a catalytically competent conformation until tRNA is bound, indicating that a set of mutually induced conformational changes occur between PaKEOPS and tRNA.

Therefore, we propose that KEOPS and DEZ must undergo conformational changes to attain a catalytically relevant conformation and that ATP hydrolysis is needed to compensate for the free energy cost of rearrangements. Further experimental work is needed to verify this hypothesis.

#### Possible role of archaeal Sua5 and KEOPS in the synthesis of N<sup>6</sup>-hydroxynorvalylcarbamoyladenine (hn<sup>6</sup>A)

Two other modified adenosines occur at position 37 and contain carbamoyl and amino acid moieties: N<sup>6</sup>-glycylcarbamoyladenine (g<sup>6</sup>A) and N<sup>6</sup>-hydroxynorvalylcarbamoyladenine (hn<sup>6</sup>A) (3,40), but the enzymes catalysing the synthesis of these modified nucleosides have not been identified yet. g<sup>6</sup>A is present in certain cytoplasmic tRNAs of eukaryotes, whereas hn<sup>6</sup>A is found in tRNAs of Bacteria and Archaea (41). In agreement with this phylogenetic distribution, we found that PaSua5 hydrolyses ATP to AMP in presence of hydroxynorvaline, but not in presence of glycine (Supplementary Figure S4B). This suggested that Sua5 proteins could be responsible for the synthesis of the hydroxynorvalylcarbamoyladenylyl and that, together with KEOPS, they may be responsible for the formation of hn<sup>6</sup>A in Archaea. Further experimental work is needed to confirm this hypothesis, which if true, would extend the role of archaeal KEOPS and Sua5 proteins to the synthesis of another tRNA modification.

#### Evolutionary implications

The present work and the previous work on bacterial system (15) revealed that the universal proteins Sua5/YrdC and KaeI/YgjD alone are not sufficient to perform t<sup>6</sup>A modification but require two sets of non-universal proteins, one specific to Bacteria (YeaZ, YjeE) and the other common to Archaea and Eukarya (Bud32, Cgi121 and Pcc1). A similar situation is found in many other fundamental molecular systems, e.g. translation initiation and transcription, in which a core of universally

conserved proteins is supplemented by factors specific for Bacteria on one side and for Archaea/Eukarya on the other. It is often argued that such a situation suggests a simple LUCA [a progenote *sensu* (42) and subsequent independent refinement of molecular processes (i.e. translation)] in Bacteria and in Archaea/Eukarya. In case of t<sup>6</sup>A, it seems unlikely that LUCA lacked this modification. Indeed, several studies identified t<sup>6</sup>A as a part of a minimal tRNA modifications set required for life (43–45). Moreover, the genes responsible for the synthesis of t<sup>6</sup>A are resistant to loss in insect symbionts and mollicutes that have reduced genomes, suggesting that these genes emerged early during cellular evolution, and once acquired became essential for the cell (45,46). Finally, it has been argued convincingly that t<sup>6</sup>A modification is necessary for the accurate reading of the three letters genetic code (7,8,19). Because this code is universal, it was likely present in LUCA, and its accuracy was probably already optimal (47).

If t<sup>6</sup>A was already present in LUCA, one should imagine evolutionary scenarios to explain how an ancestral tRNA modification mechanism evolved into two distinct modern versions. The idea that these two versions evolved from a common mechanism is supported by the existence of two universal proteins in both versions, by the universality of the two distinct ATP-dependent steps and by the interdomain functional complementations. One possible scenario is that the core proteins (Sua5/YrdC and KaeI/YgjD) are the remnants of an ancient synthetase operating in LUCA, which could have performed the complete reaction without accessory proteins. After the separation of the three domains, additional proteins were added independently in Bacteria and in ancestors of Archaea and Eukarya to improve speed and accuracy of protein translation in increasingly complex cells. In this scenario, we propose that these proteins were selected to provide additional surfaces for tRNA binding. This suggestion fits well with the observation that KaeI alone binds tRNA less tightly as compared with KEOPS (as explained previously). Alternatively, LUCA might have possessed either DEZ or KEOPS to work with the ancestor of Sua5/YrdC; later on, YeaZ and YjeE were replaced by Bud32, Pcc1 and Cgi121 in Archaea and Eukarya or *vice versa*, and KEOPS was replaced by DEZ in Bacteria. Our complementation experiments render this hypothesis plausible because we show that both KEOPS and DEZ can complement *in vitro* the Sua5/YrdC protein of the three different domains, suggesting that such situation could have happened *in vivo* in ancestors of modern organisms.

The study of t<sup>6</sup>A synthesis in mitochondria could contribute to the comprehension of t<sup>6</sup>A machinery evolution. Eukaryotic genomes encode mitochondrial homologue of KaeI called Qri7 in *S. cerevisiae*. Genetic data from two independent studies indicated that *S. cerevisiae* Qri7 functions in mitochondrial t<sup>6</sup>A synthesis. Interestingly, Sua5 is required for both cytoplasmic and mitochondrial t<sup>6</sup>A synthesis, indicating that it is imported into mitochondria despite the lack of a recognizable import signal (14,20). It remains to be investigated if Qri7 and Sua5 alone could catalyse the formation of t<sup>6</sup>A-modified tRNA or if other



proteins are required. In the case of the second hypothesis, the proteins of KEOPS are the obvious candidates. The first hypothesis, if true, would substantially argue in favour of the scenario in which the two universal proteins formed the ancient t<sup>6</sup>A synthetic machinery.

In conclusion, this study established an *in vitro* system for the synthesis of essential t<sup>6</sup>A tRNA modification by eukaryotic and archaeal t<sup>6</sup>A synthetases. Characterization of these systems allowed us to propose a novel reaction scheme for the synthesis of t<sup>6</sup>A. Our data suggest that the catalytic mechanism for t<sup>6</sup>A synthesis is conserved in nature. Our work paves the way for further mechanistic and structural studies of these enzymatic complexes. The full comprehension of how this ancient modification is synthesized should result in better understanding of evolutionary mechanisms that contributed to increased translation efficiency and fidelity in present day cells.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–7, Supplementary Methods.

## ACKNOWLEDGEMENTS

The authors are indebted to Henri Grosjean for his encouragement, stimulating discussions and critical reading of the manuscript. The authors thank Domenico Libri and Marie-Claire Dageron for helpful discussions and for providing the Sua5 yeast deletion strain; Dorian Guetta, Philippe Bouloc, Chantal Bohn and Emmanuelle Schmitt for sharing protocols and for helpful suggestions. B.C. wishes to warmly thank Magali Bland for her helpful advices and protocols and François Mallet for his generous gift of pMRT-7 wt.

## FUNDING

Agence Nationale de Recherche [ANR-09-BLAN-0349 to P.F. and H.v.T.]; Marie Curie postdoctoral fellowship (to E.C.); ENS Lyon (PhD fellowship to L.P.). Funding for open access charge: Agence Nationale de Recherche [ANR-09-BLAN-0349].

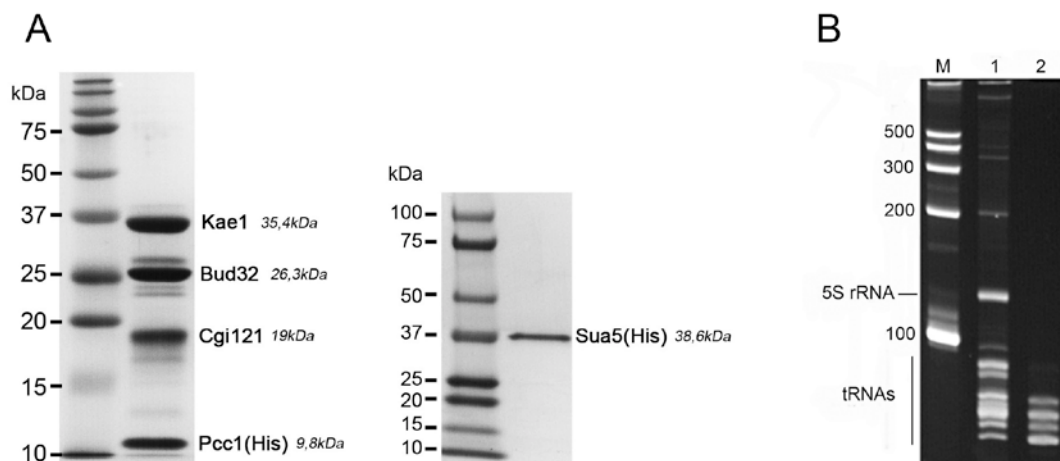
*Conflict of interest statement.* None declared.

## REFERENCES

- Grosjean, H., Sprinzl, M. and Steinberg, S. (1995) Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie*, **77**, 139–141.
- Gustilo, E.M., Vendeix, F.A. and Agris, P.F. (2008) tRNA's modifications bring order to gene expression. *Curr. Opin. Microbiol.*, **11**, 134–140.
- Grosjean, H. Nucleic acids are not boring long polymers of only four types of nucleotides. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Biosciences, Austin, Texas, USA, pp. 1–18.
- Juhling, F., Morl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F. and Putz, J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.*, **37**, D159–D162.
- Giegé, R., Juhling, F., Pütz, J., Stadler, P., Sauter, C. and Florentz, C. (2012) Structure of transfer RNAs: similarity and variability. *Wiley Interdiscipl. Rev. RNA*, **3**, 37–61.
- Auffinger, P. and Westhof, E. Location and distribution of modified nucleotides in tRNA. In: Grosjean, H. and Benne, R. (eds), *Modification and editing of RNA*. ASM press, Washington DC, USA, pp. 103–112.
- Stuart, J.W., Gdaniec, Z., Guenther, R., Marszałek, M., Sochacka, E., Malkiewicz, A. and Agris, P.F. (2000) Functional anticodon architecture of human tRNA<sup>Lys3</sup> includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t<sup>6</sup>A. *Biochemistry*, **39**, 13396–13404.
- Murphy, F.V., Ramakrishnan, V., Malkiewicz, A. and Agris, P.F. (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.*, **11**, 1186–1191.
- Chheda, G.B., Hall, R.H., Mozejko, J., Magrath, D.I., Schweizer, M.P., Stasiuk, L. and Taylor, P.R. (1969) Aminoacyl nucleosides. VI. Isolation and preliminary characterization of threonyladenine derivatives from transfer ribonucleic acid. *Biochemistry*, **8**, 3278–3282.
- Chheda, G., Hong, C. and Piskorz, C. (1972) Biosynthesis of N-(purin-6-ylcarbamoyl)-L-threonine riboside. Incorporation of L-threonine in vivo into modified nucleoside of transfer ribonucleic acid. *Biochem. J.*, **127**, 515–519.
- Körner, A. and Söll, D. (1974) N-(purin-6-ylcarbamoyl)threonine: biosynthesis in vitro in transfer RNA by an enzyme purified from *Escherichia coli*. *FEBS Lett.*, **39**, 301–306.
- Galperin, M.Y. (2004) Conserved hypothetical proteins: prioritization of targets for experimental study. *Nucleic Acids Res.*, **32**, 5452–5463.
- El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J.R., Schimmel, P., Swairjo, M.A. and de Crecy-Lagard, V. (2009) The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA. *Nucleic Acids Res.*, **37**, 2894–2909.
- El Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J., Iwata-Reuyl, D., Murzin, A.G. and de Crecy-Lagard, V. (2011) A role for the universal Kae1/Qri7/YgiD (COG0533) family in tRNA modification. *EMBO J.*, **30**, 882–893.
- Deutsch, C., El Yacoubi, B., de Crecy-Lagard, V. and Iwata-Reuyl, D. (2012) Biosynthesis of Threonylcarbamoyl Adenosine (t<sup>6</sup>A), a Universal tRNA Nucleoside. *J. Biol. Chem.*, **287**, 13666–13673.
- Lauhon, C. (2012) Mechanism of N<sup>6</sup>-threonylcarbamoyladenine (t<sup>6</sup>A) biosynthesis: isolation and characterization of the intermediate threonylcarbamoyl-AMP. *Biochemistry*, **51**, 8950–8963.
- Downey, M., Houlsworth, R., Maringe, L., Rollic, A., Brehme, M., Galicia, S., Guillard, S., Partington, M., Zubko, M.K., Krogan, N.J. et al. (2006) A Genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell*, **124**, 1155–1168.
- Kisseleva-Romanova, E., Lopreiato, R., Baudin-Baillieu, A., Rousselle, J., Ilan, L., Hofmann, K., Namane, A., Mann, C. and Libri, D. (2006) Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO J.*, **25**, 3576–3585.
- Dageron, M., Lenstra, T.L., Frizzarin, M., El Yacoubi, B., Liu, X., Baudin-Baillieu, A., Lijnzaad, P., Decourty, L., Saveanu, C., Jacquier, A. et al. (2011) Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t<sup>6</sup>A modification of tRNAs. *Nucleic Acids Res.*, **39**, 1–13.
- Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, V.E., Karzai, A.W. and Sternglanz, R. (2011) The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t<sup>6</sup>A. *EMBO J.*, **30**, 873–881.
- Naor, A., Thiaville, P.C., Altman-Price, N., Cohen-Or, I., Allers, T., de Crécy-Lagard, V. and Gophna, U. (2012) A genetic investigation of the KEOPS complex in halophilic Archaea. *PLoS One*, **7**, e43013.
- Abdullah, K.M., Lo, R.Y. and Mellors, A. (1991) Cloning, nucleotide sequence, and expression of the *Pasteurella haemolytica* A1 glycoprotease gene. *J. Bacteriol.*, **173**, 5597–5603.
- Collinet, B., Friberg, A., Brooks, M.A., Elzen, T., Henriot, V., Dziembowski, A., Graille, M., Dur, D., Leulliot, N., Saint

- André, C.S. *et al.* (2011) Strategies for the structural analysis of multi-protein complexes: lessons from the 3D-Repertoire project. *J. Struct. Biol.*, **175**, 147–158.
24. Hecker, A., Leulliot, N., Gabelle, D., Graille, M., Justome, A., Dorlet, P., Brochier, C., Quevillon-Cheruel, S., Le Cam, E., Le Cam, E. *et al.* (2007) An archaeal orthologue of the universal protein Kae1 is an iron metalloprotein which exhibits atypical DNA-binding properties and apurinic-endonuclease activity in vitro. *Nucleic Acids Res.*, **35**, 6042–6051.
25. Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
26. Mougin, A., Grégoire, A., Banroques, J., Ségault, V., Fournier, R., Brulé, F., Chevrier-Miller, M. and Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079–1093.
27. Arnaud, N., Cheynet, V., Oriol, G., Mandrand, B. and Mallet, F. (1997) Construction and expression of a modular gene encoding bacteriophage T7 RNA polymerase. *Gene*, **199**, 149–156.
28. Grosjean, H., Droogmans, L., Roovers, M. and Keith, G. (2007) Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods Enzymol.*, **425**, 55–101.
29. Mao, D.Y.L., Neculai, D., Downey, M., Orlicky, S., Haffani, Y.Z., Ceccarelli, D.F., Ho, J.S.L., Szilard, R.K., Zhang, W., Ho, C.S. *et al.* (2008) Atomic structure of the KEOPS complex: an ancient protein kinase-containing molecular machine. *Mol. Cell*, **32**, 259–275.
30. Morin, A., Auxilien, S., Senger, B., Tewari, R. and Grosjean, H. (1998) Structural requirements for enzymatic formation of threonylcarbamoyladenine (t6A) in tRNA: an in vivo study with *Xenopus laevis* oocytes. *RNA*, **4**, 24–37.
31. Madore, E., Florentz, C., Giegé, R., Sekine, S., Yokoyama, S. and Lapointe, J. (1999) Effect of modified nucleotides on *Escherichia coli* tRNA<sup>Glu</sup> structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. *Eur. J. Biochem.*, **266**, 1128–1135.
32. Tisné, C., Rigourd, M., Marquet, R. and Ehresmann, C. (2000) NMR and biochemical characterization of recombinant human tRNA (Lys) 3 expressed in *Escherichia coli*: identification of posttranscriptional nucleotide modifications. *RNA*, **6**, 1403–1412.
33. Grosjean, H., Keith, G. and Droogmans, L. (2004) Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. *Methods Mol. Biol.*, **265**, 357–391.
34. Stocchetto, S., Marin, O., Carignani, G. and Pinna, L.A. (1997) Biochemical evidence that *Saccharomyces cerevisiae* YGR262c gene, required for normal growth, encodes a novel Ser/Thr-specific protein kinase. *FEBS Lett.*, **414**, 171–175.
35. Meng, F., Chen, X., Hu, Y., Tang, H., Dang, W. and Zhou, J. (2010) Sua5p is required for telomere recombination in *Saccharomyces cerevisiae*. *Cell Res.*, **20**, 495–498.
36. Wei, F., Suzuki, T., Watanabe, S. and Kimura, S. (2011) Deficit of tRNA<sup>Lys</sup> modification by Cdk11 causes the development of type 2 diabetes in mice. *J. Clin. Invest.*, **121**, 3598–3608.
37. Garcia, G.A. and Goodenough-Lashua, D.M. Mechanisms of RNA modifying and editing enzymes. In: Dans Grosjean, H. and Benne, R. (eds), *Modification and Editing of RNA*. ASM Press, Washington DC, pp. 135–168.
38. Kuratani, M., Kasai, T., Akasaka, R., Higashijima, K., Terada, T., Kigawa, T., Shinkai, A., Bessho, Y. and Yokoyama, S. (2011) Crystal structure of *Sulfolobus tokodaii* sua5 complexed with L-threonine and AMPPNP. *Proteins: Structure, Function, and Bioinformatics*, **79**, 2065–2075.
39. Parthier, C., Görlich, S., Jaenecke, F., Breithaupt, C., Bräuer, U., Fandrich, U., Clausnitzer, D., Wehmeier, U.F., Böttcher, C., Scheel, D. *et al.* (2012) The O-Carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction. *Angew. Chem. Int. Ed. Engl.*, **51**, 4046–4052.
40. Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D. and Agris, P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.
41. Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M., Grosjean, H. and Rother, K. (2009) MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res.*, **37**, D118–D121.
42. Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA*, **74**, 5088–5090.
43. Björk, G.R., Ericson, J.U., Gustafsson, C.E., Hagervall, T.G., Jönsson, Y.H. and Wikström, P.M. (1987) Transfer RNA modification. *Annu. Rev. Biochem.*, **56**, 263–287.
44. Forster, A.C. and Church, G.M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2**, doi: 10.1038/msb4100090.
45. de Crecy-Lagarde, V., Marck, C. and Grosjean, H. (2012) Decoding in *Candidatus Riesia pediculicola*, close to minimal tRNA modification set? *Trends Cell Mol. Biol.*, **7**, 11–34.
46. de Crecy-Lagarde, V., Marck, C., Brochier-Armanet, C. and Grosjean, H. (2007) Comparative RNomics and modomics in mollicutes: prediction of gene function and evolutionary implications. *IUBMB Life*, **59**, 634–658.
47. Vetsigian, K., Woese, C. and Goldenfeld, N. (2006) Collective evolution and the genetic code. *Proc. Natl Acad. Sci. USA*, **103**, 10696–10701.

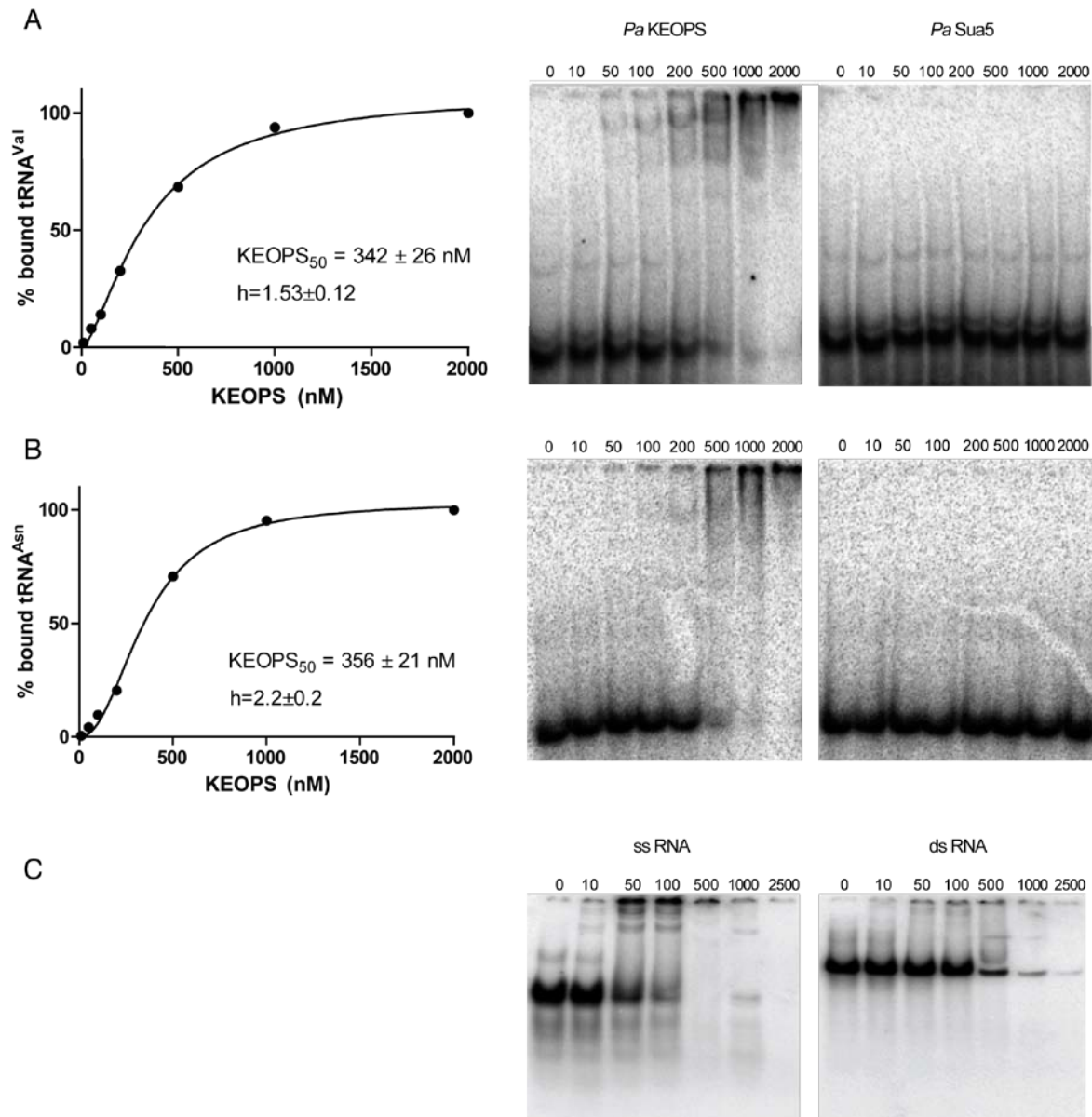
## SUPPLEMENTARY DATA



**Figure S1. Purification of *P. abyssi* KEOPS complex and Sua5.**

(A) Co-purification of KEOPS subunits. SDS-PAGE analysis of purified KEOPS from *P. abyssi*. The different subunits are indicated, along with their molecular mass. Only Pcc1 carries a C-terminal polyhistidine tag. Sua5 was purified independently and carried a C-terminal hexahistidine or Strep-tag.

(B) Denaturing polyacrylamide gel analysis of nucleic acid bands co-purifying with PaKEOPS. (M), ssRNA size marker; (1), total RNAs extracted from *E. coli* strain used for expression of PaKEOPS; (2) RNAs co-purifying with KEOPS. The sizes in nucleotides and identity of the main bands are indicated on the left.



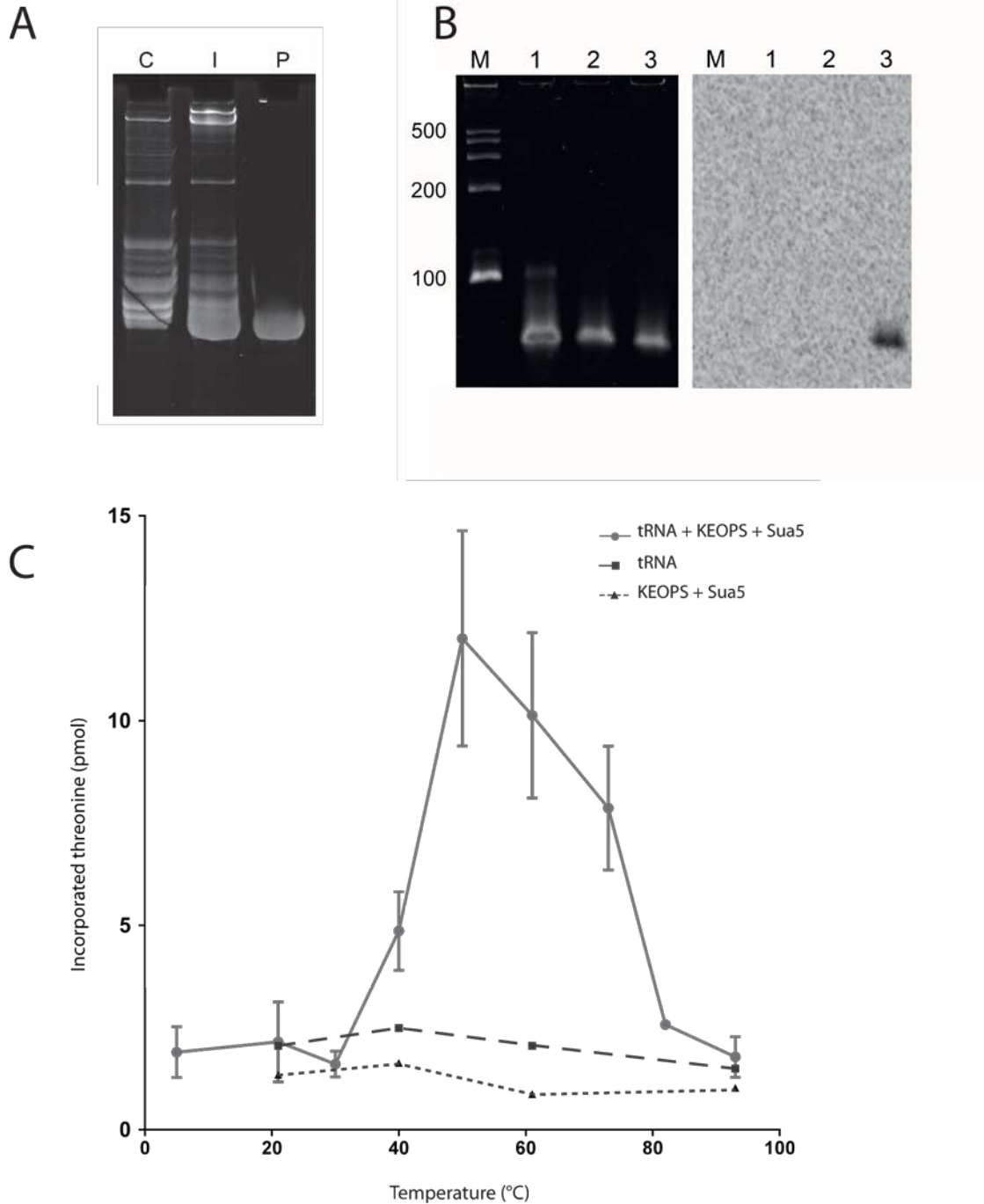
**Figure S2. Binding of PaKEOPS and PaSua5 to RNAs.**

Electromobility shift assays showing binding of PaKEOPS (left) or PaSua5 (right) to radiolabelled Ec\_tRNA<sup>Val</sup> (A - non ANN codon), Ec\_tRNA<sup>Asn</sup> (B – ANN codon). The concentration of proteins used (nM) in each reaction is indicated on top of each well.

The percent of bound tRNA was plotted as a function of KEOPS concentration and the data were fitted using nonlinear regression least squares fit for one site specific binding model with Hill plot. The concentration of PaKEOPS at which 50 % of tRNA was bound is indicated.

The data were best fitted ( $R^2 > 0.99$ ) using one site specific cooperative binding model, with the derived hill constant of  $2.2 \pm 0.2$  for tRNA<sup>Asn</sup> and  $1.53 \pm 0.12$  for tRNA<sup>Val</sup>, indicating positive cooperativity for tRNA binding by KEOPS. The calculated dissociation constants were  $342 \pm 26$  nM for tRNA<sup>Val</sup> and  $356 \pm 21$  nM for tRNA<sup>Asn</sup>.

(C) Binding of PaKEOPS on a random 65-mer ssRNA (left panel) or dsRNA (right panel).



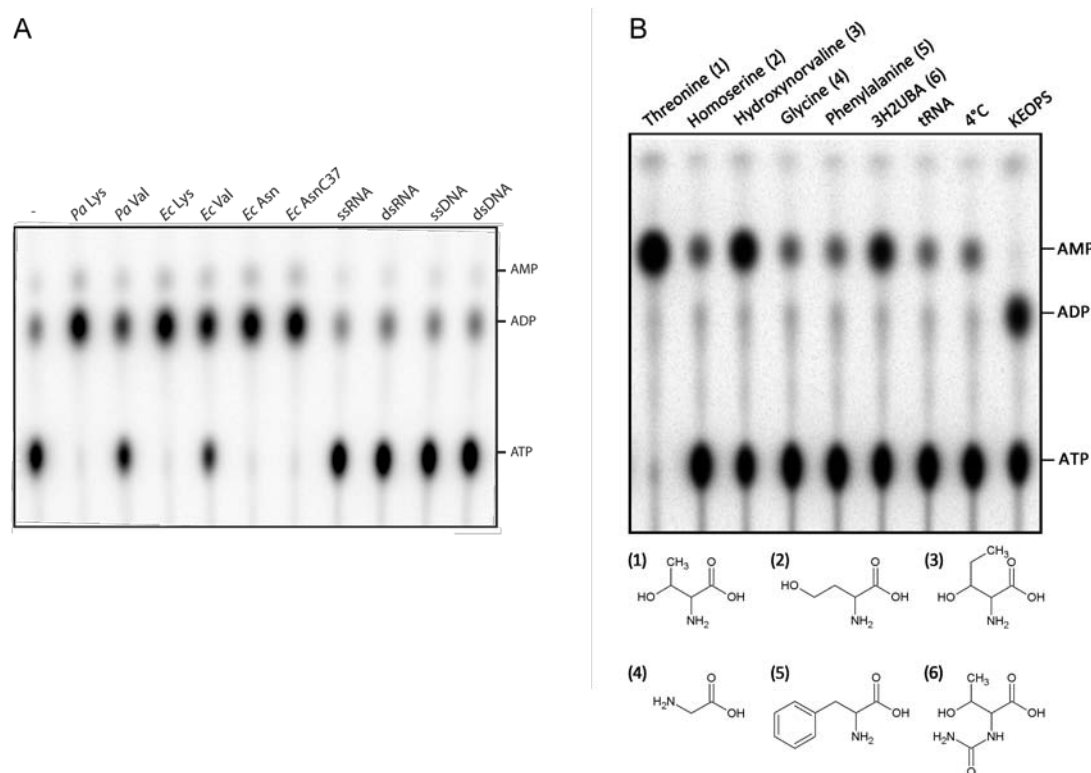
**Figure S3. Details of  $t^6A$  *in vitro* reaction.**

(A) Heterologous expression and purification of Pa\_tRNA<sup>Lys</sup>. The gene encoding Pa\_tRNA<sup>Lys</sup> was overexpressed in *E. coli* cells under the control of IPTG inducible promoter. Total RNA was isolated from induced cells (lane I) and non-induced control cells (lane C). The band corresponding to the purified Pa\_tRNA<sup>Lys</sup> (lane P).

(B) ATP dependent incorporation of radiolabelled threonine into Pa\_tRNA<sup>Lys</sup>. The runoff transcripts of Pa\_tRNA<sup>Lys</sup> (lane 1) were used as substrate for  $t^6A$  assay in absence (lane 2) or presence (lane 3) of ATP. tRNA was subsequently isolated from the reaction mixture and separated on polyacrylamide gel

(left panel). The radioactivity was visualized by phosphorimaging. (M), ssRNA molecular size marker; sizes in nucleotides are indicated.

(C) Formation of  $t^6A$  by PaKEOPS and PaSua5 as a function of temperature. Incorporation of radioactive threonine in TCA precipitated material was measured at different temperatures. In control experiments PaSua5 and PaKEOPS (squares) or tRNA substrate (triangles) were omitted. Error bars indicate deviation from the mean for two independent experiments.

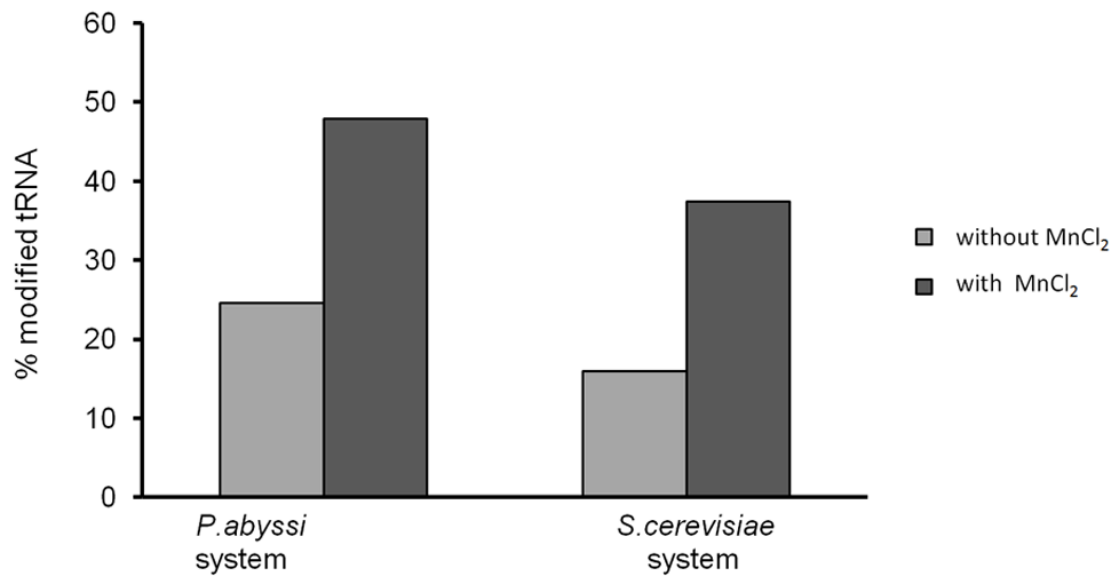


**Figure S4. Activation of KEOPS and Sua5 ATPase activity.**

(A) ATPase activity of PaKEOPS in presence of tRNA or random sequences. KEOPS was incubated with  $\alpha$ - $^{32}P$ -ATP and different tRNAs or random sequence 65mer of single or double stranded RNA or DNA. The control experiment (-) was performed at 50°C but the RNA was omitted.

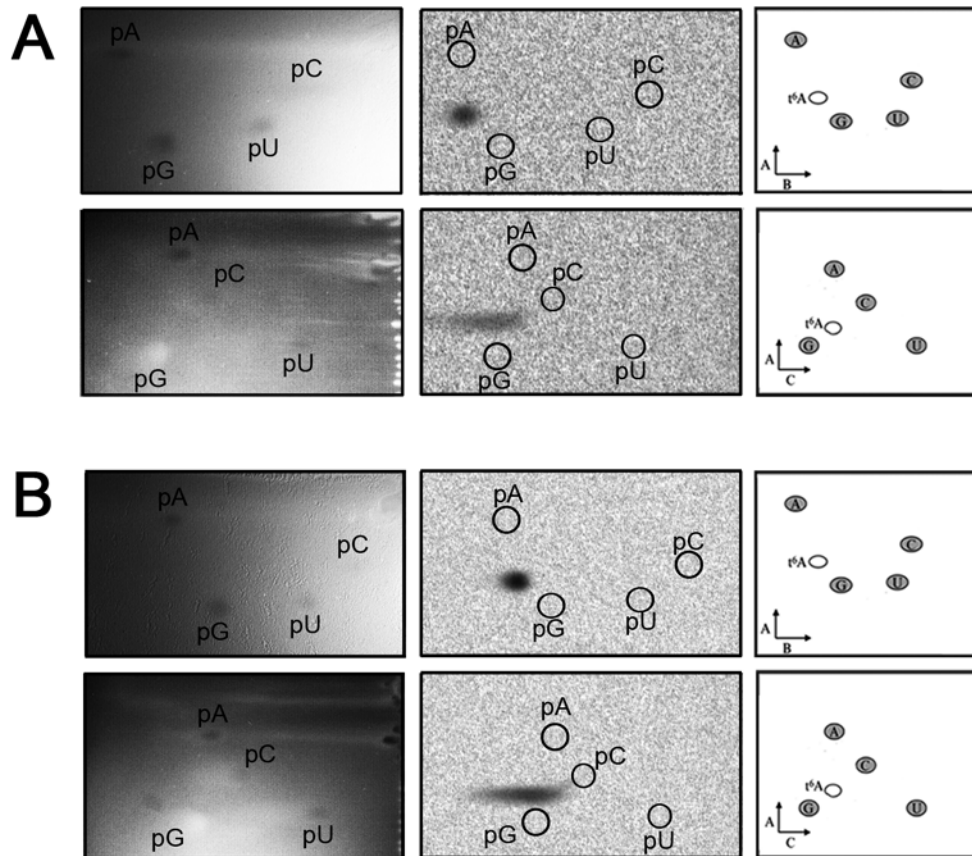
(B) ATPase activity of PaSua5 in presence of different amino-acids. PaSua5 was incubated with  $\alpha$ - $^{32}P$ -ATP and different amino-acids as substrates, as indicated above each reaction. 3H2UBA is 3-Hydroxy-2-ureido-butyric acid. The control experiments were performed with tRNA (tRNA), without any substrate (4°C), or with PaKEOPS incubated with tRNA to locate ADP on the plate (KEOPS). Below, chemical structures of the different compounds are shown.





**Figure S5. Effect of manganese ions on t<sup>6</sup>A synthesis.**

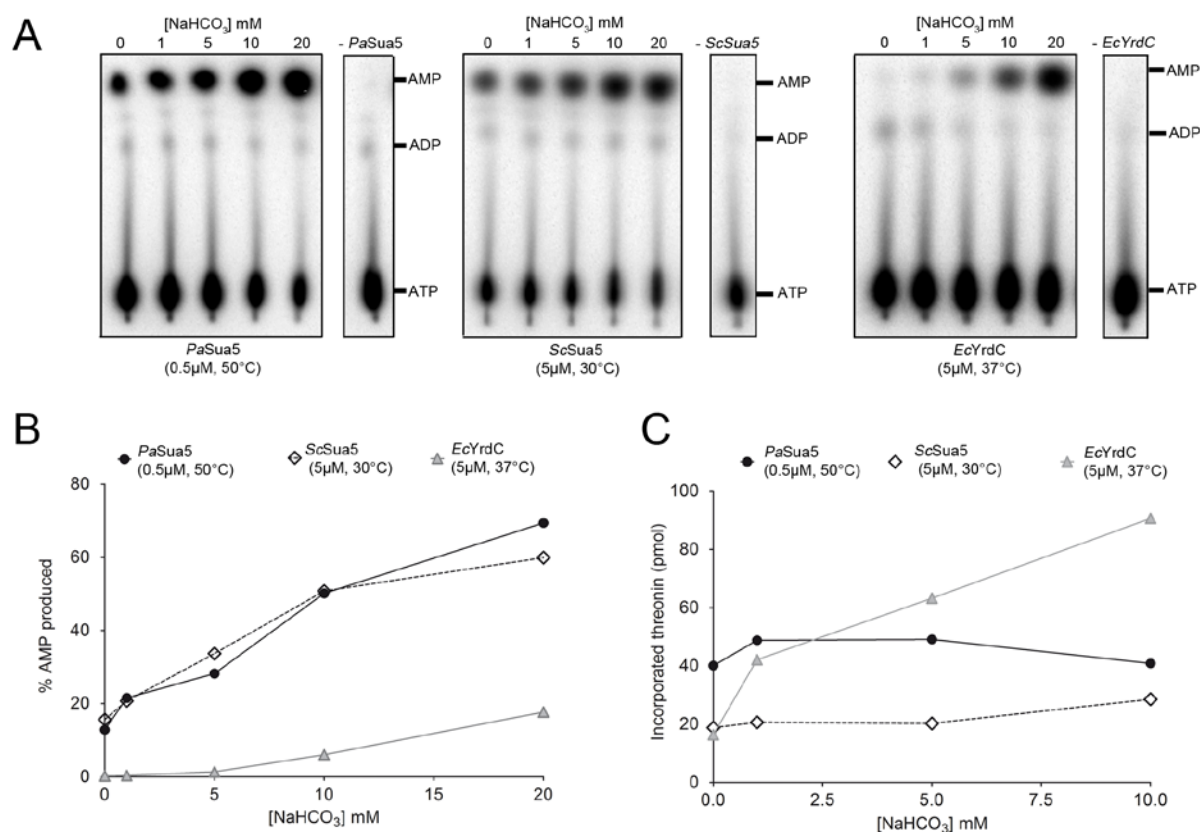
t<sup>6</sup>A assay was performed in presence or absence of manganese divalent ions for *P. abyssi*, *S. cerevisiae* and *E. coli* systems and the incorporation of radioactive threonine in TCA precipitated material was measured as described in materials and methods. Ec\_tRNAAsn produced *in vitro* was used as substrate.



**Figure S6. 2D TLC analysis of  $t^6A$  assay performed with archaeal (A) and mixed bacterial/archaeal system (B).**

Upper line shows chromatographic mobility in solvent system A+B, lower line in solvent system A+C. Left panels: abundant, non-modified nucleotides were detected by UV shadowing. Center panel:  $^{14}C$ -threonine incorporated in modified nucleotides was detected by phosphorimaging, and the positions of unmodified nucleotides were superimposed by aligning with the UV image (white circles). Right panel: reference maps (from (1)) showing the theoretical position of  $t^6A$  nucleoside in both solvent systems.  $t^6A$  assay was performed at 50 °C for archaeal system or at 43°C for mixed system obtained by combining *E. coli* DEZ proteins and *P. abyssi* Sua5. Ec\_tRNA<sup>Lys</sup> produced in overexpressing *E. coli* strain was used as a substrate.





**Figure S7. Effect of carbonate on  $t^6A$  synthesis.**

(A) Effect of carbonate on ATPase activity of PaSua5, ScSua5 and EcYrdC proteins. Produced radiolabelled ( $\alpha$ -<sup>32</sup>P) ADP and AMP were separated by TLC and visualised by phosphorimaging. The final concentration of added carbonate is indicated above each well. Pa\_tRNA<sup>Lys</sup> produced in overexpressing *E. coli* strain was used as a substrate.

(B) ATPase activity of PaSua5, ScSua5 and EcYrdC proteins as a function of added carbonate concentration. The radioactive spots corresponding to AMP shown in part (A) were quantified using ImageQuant. Incubation time and temperature for each protein are indicated. The protein concentrations used were as follows: 0.5  $\mu$ M for PaSua5, 5  $\mu$ M for ScSua5 and EcYrdC.

(C) Effect of carbonate on total reaction leading to  $t^6A$  synthesis. Threonine incorporation in TCA precipitated material was measured at different concentration of added carbonate using *P. abyssi*, *S. cerevisiae* or *E. coli*  $t^6A$  synthetase. Incubation time and temperature for each system are indicated.

### **Buffers composition**

**Lysis Buffer (LBf):** 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM  $\beta$ -mercaptoethanol, 10mM imidazole, 5% glycerol

**Transcription Buffer (TBf):** 40 mM Tris-HCl pH 8, 20 mM  $MgCl_2$ , 5 mM DTT, 1 mM spermidine, 0,01 % triton and 80 mg/mL PEG 8000 supplemented with 200 nmol of each rNTP, 40 U of RNasine (Thermo Scientific), 300 ng of PCR product and 0.2  $\mu$ M of his-T7 RNA polymerase expressed and purified according to (2).

**Elution Buffer (EB) :** 0,1% SDS, 0.5 M  $NH_4Ac$ , 10 mM  $MgAc$ , 1 mM EDTA

**Reaction Buffer (RB):** 50 mM Tris-HCl pH8, 35 mM KCl, 10 mM  $MgCl_2$  and 1 mM  $MnCl_2$ , 5 mM DTT, 1 mM ATP, 10 mM  $NaHCO_3^-$ .

### **Chromatography solvents (1):**

Solvent A: isobutyric acid/concentrated ammonia/water; 66:1:33 [v/v/v]

Solvent B: phosphate buffer/ammonium sulfate/n-propanol; 100:60:2 [v/w/v]

Solvent C: isopropanol/concentrated HCl/water; 68:18:14 [v/v/v]

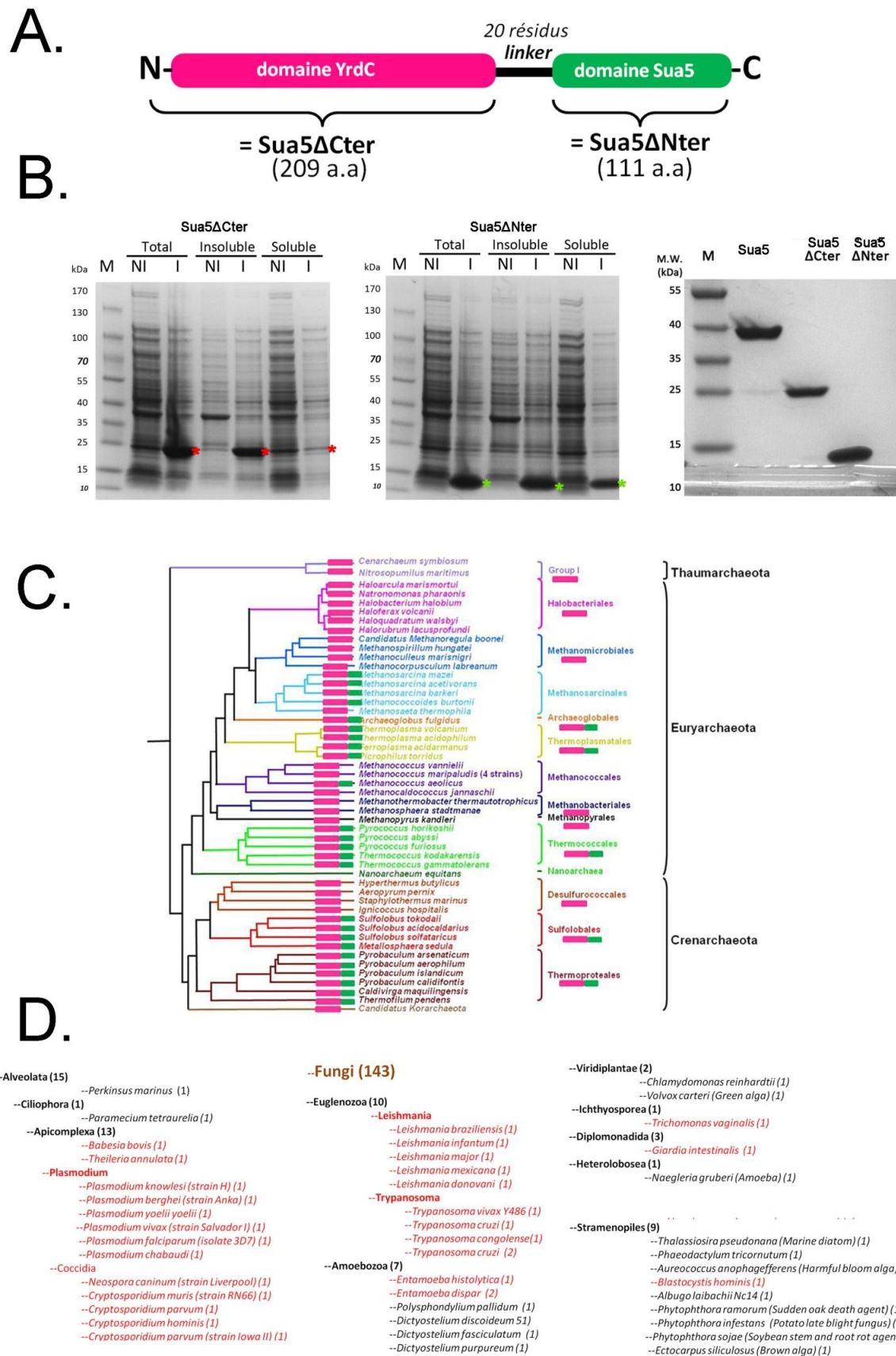
**Binding Buffer (BB) :** 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1mM DTT

**TGE Buffer:** 50 mM Tris-HCl pH 7.5, 8 mM glycine, 0.1 mM EDTA

### **REFERENCES**

1. Grosjean,H., Droogmans,L., Roovers,M. et Keith,G. (2007) Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods in enzymology*, **425**, 55-101.
2. Arnaud,N., Cheynet,V., Oriol,G., Mandrand,B. et Mallet,F. (1997) Construction and expression of a modular gene encoding bacteriophage T7 RNA polymerase. *Gene*, **199**, 149-56.





**FIGURE I-1 : Etude des domaines N et C terminaux de Sua5 (suite de la légende page suivante)**

**A. Représentation schématique des domaines protéiques de la protéine Sua5 de *Pyrococcus abyssi***

Les protéines Sua5ΔCter et Sua5ΔNter indiquées, correspondant respectivement aux domaines YRDC et SUA5, ont été produites sous forme de protéines recombinantes fusionnées avec une étiquette StrepTagII en position C-terminale.

## Résultats complémentaires :

Les résultats présentés par la suite accompagnent et complètent les travaux précédents. Ils n'ont pas été inclus dans la publication présentée car ils constituent des données préliminaires, mais ils fournissent des pistes intéressantes dans la compréhension de la machinerie de synthèse de t<sup>6</sup>A chez *Pyrococcus abyssi*.

### Rôle du domaine C-terminale de Sua5 :

Nous avons montré que la protéine Sua5 de *P.abysssi* est indispensable à la biosynthèse de t<sup>6</sup>A, et que les orthologues de *S.cerevisiae* et *E.coli* présentent des activités semblables et sont interchangeables. Or, les protéines Sua5 de *P.abysssi* et de *S.cerevisiae* présentent un domaine C-terminal d'environ 100 acides aminés (annoté « SUA5 ») que ne possède pas l'orthologue YrdC de *E.coli* (FIGURE I-1-A). Ce domaine C-terminal est lié au domaine YrdC par un *linker* d'environ 20 résidus. Afin de savoir si la présence de domaine spécifique est nécessaire à l'activité de la protéine Sua5 chez *P.abysssi*, nous avons alors exprimé séparément le domaine N-terminal (YrdC) et le domaine C-terminal (SUA5), chacun étant fusionné avec une étiquette StrepTag-II en position C-terminale. Ces deux protéines, appelées respectivement Sua5ΔCter et Sua5ΔNter, ont pu être sur-exprimées chez *E.coli*. Après la lyse cellulaire, nous avons observé que Sua5ΔCter se retrouve en majorité dans la fraction insoluble du lysat cellulaire, alors que Sua5ΔNter se retrouve à la fois dans les fractions solubles et insolubles en quantité approximativement égale, (FIGURE I-1B), comme la protéine complète Sua5. Cela suggère que le domaine C-terminale pourrait jouer un rôle dans la solubilisation du domaine N terminale. Nous avons néanmoins réussi à purifier jusqu'à homogénéité chacun de ces deux domaines par chromatographie d'affinité suivie d'une filtration sur gel (FIGURE I-1B).

La protéine Sua5ΔCter, correspondant au domaine catalytique YrdC, n'est plus capable de catalyser la biosynthèse de t<sup>6</sup>A en présence de KEOPS. L'ajout de Sua5ΔNter à la réaction n'a aucun effet (non montré). De même, les deux protéines séparées ou combinées ne présentent pas d'activité d'hydrolyse de l'ATP en AMP, contrairement à la protéine entière Sua5 (non montré). Nous avons alors voulu vérifier si les deux domaines purifiés séparément étaient capables d'interagir. Nous avons incubé les protéines Sua5ΔCter et Sua5ΔNter dans différentes conditions de température et de temps d'incubation, puis le

## FIGURE I-1 : suite de la légende (p172)

### B. Expression et purification des protéines tronquées Sua5ΔCter et Sua5ΔNter

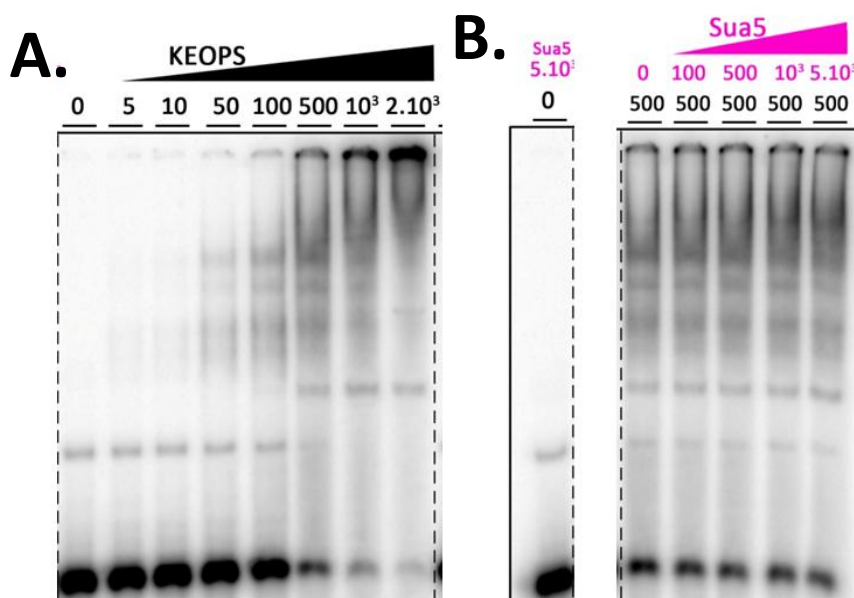
Les protéines Sua5ΔCter et Sua5ΔNter ont été sur-exprimées dans une souche *Escherichia coli* Rosetta2(DE3)pLysS en condition inductrice (I). En parallèle, une culture identique a été réalisée en condition non inductrice (NI). Après lyse des cellules, l'extrait cellulaire total a été centrifugé à 10.000g pendant 30mn pour séparer la fraction insoluble de la fraction soluble. Pour chacune des protéines, les différentes fractions ont été analysées en SDS PAGE. Les étoiles rouges et vertes indiquent la présence des protéines recombinantes. La fraction soluble contenant chacune des protéines a été purifiée selon le matériel et méthode, puis les protéines purifiées (2μL) ont été analysées par SDS PAGE.

### C. Distribution du domaine SUA5-Cterminal chez les Archées selon l'arbre phylogénétique

L'arbre phylogénétique des Archées représenté est basé sur l'analyse des protéines ribosomales (Brochier *et al.* 2008). L'orthologue de Sua5 pour chaque espèce représentative a été identifié par Blastp, et a été représenté en face de chaque branche. Le rectangle rose correspond au domaine YrdC, le rectangle vert correspond au domaine SUA5. Lorsque toutes les espèces d'un même phylum présentent un orthologue similaire (YrdC seul ou YrdC-Sua5), celui-ci a été représenté sous le nom du phylum correspondant.

### D. Distribution du domaine SUA5-Cterminal chez les Eucaryotes

La distribution du domaine SUA5 (PF03481) chez les Eucaryotes a été déterminée par la banque de données protéiques Pfam version 27.1 for PF03481. Les différents niveaux de classification ont été simplifiés pour plus de clarté. Le phylum des *Fungi* a été mis en évidence en marron. Les espèces parasites de l'Homme et des mammifères sont indiquées en rouge. Le nombre entre parenthèse à côté de chaque groupe correspond au nombre d'espèces présentant un exemplaire du domaine SUA5.



## FIGURE I-2 : Retard sur gel des ARNt par KEOPS en présence de Sua5

### A. Fixation de KEOPS à l'ARNt.

Une concentration fixe (10nM) d'ARNt Lys(UUU) de *P.abyssi* sur-exprimés *in vivo* et radiomarké a été incubée en présence de différentes concentrations de complexe KEOPS, dans un tampon TrisHCl 50mM, pH8, NaCl 100mM, glycérol 5%, MgCl 5mM. Les complexes nucléo-protéiques ont été séparés par électrophorèse sur gel d'acrylamide 6% en condition native. La concentration de KEOPS est indiquée en nM.

### B. Fixation de KEOPS à l'ARNt en présence de Sua5

La colonne au centre est une condition contrôle en l'absence de KEOPS et en présence de 5μM de Sua5. Dans l'encart de droite, une concentration fixe d'ARNt (10nM) et de KEOPS (500nM) ont été incubés en présence de différentes concentrations de Sua5 (indiqué en rose, en nM).

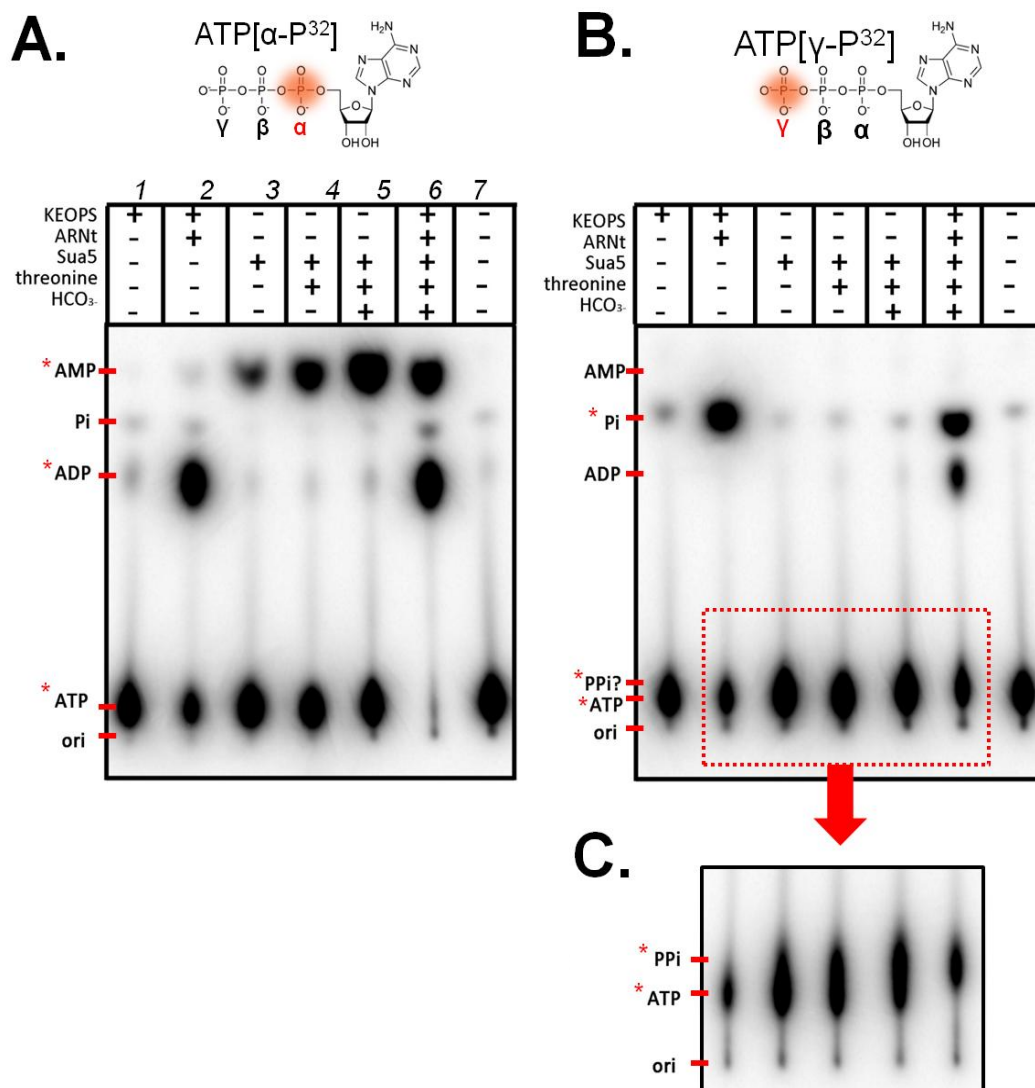
mélange a été séparé par filtration sur gel. Dans tous les cas, les deux domaines éluent sous la forme de deux pics séparés, indiquant qu'aucune interaction n'intervient dans les conditions testées (non montré).

Enfin, nous avons analysé la distribution de ce domaine SUA5 chez les Archées et chez les Eucaryotes en consultant les banques de données protéiques (PFAM). Ce domaine est retrouvé uniquement en position C-terminal du domaine YrdC, et nous n'avons pas détecté ce domaine de manière isolée ou fusionné à d'autres protéines. Ce domaine n'est présent que chez certains groupes d'Archées et sa distribution ne suit pas la phylogénie établie (**FIGURE I-1C**). Chez les Eucaryotes (**FIGURE I-1C**), ce domaine n'est également présent que chez certaines espèces. Sur les 192 espèces Eucaryotes où l'on retrouve ce domaine, la majorité fait partie des *Fungi* (143), dont *S.cerevisiae*. De manière intéressante, de très nombreux Eucaryotes unicellulaires parasites de l'Homme ou des mammifères (en rouge) présentent ce domaine, dont *Trypanosoma* (responsable de la maladie du sommeil), *Plasmodium* (responsable du paludisme) et *Leishmania* (agent de la leishmaniose).

### Une interaction directe entre Sua5 et KEOPS?

Le complexe KEOPS constituant avec Sua5 l'autre élément nécessaire à la biosynthèse de  $t^6A$ , nous avons cherché à mettre en évidence une interaction directe entre ces deux composants. Nous avons purifié la protéine Sua5 sous forme d'une protéine recombinante portant une étiquette StrepTag-II en position C-terminale. Nous avons alors testé l'interaction entre Sua5 et le complexe KEOPS (dont la protéine Pcc1 porte une étiquette hexhistidine), et entre Sua5 et la protéine Kae1-(His) par purification d'affinité et par filtration sur gel. Après incubation de ces protéines dans différentes conditions (incubation à diverses températures, en présence des différents substrats de la réaction  $t^6A$ ), nous n'avons pas pu mettre en évidence d'interaction entre Sua5 et Kae1 ou entre Sua5 et KEOPS (non montré).

N'ayant pu obtenir d'interaction physique entre Sua5 et KEOPS par ces méthodes, nous avons tenté de déterminer si l'ARNt pourrait constituer une plate-forme d'interaction entre ces deux composants. En effet, nous avons montré précédemment que KEOPS se lie aux ARNt et provoque un retard de migration sur gel (**FIGURE I-2A**), indiquant la formation d'un complexe KEOPS/ARNt stable, alors que la protéine Sua5 seule ne présente pas cette



### FIGURE I-3 : Suivi parallèle de l'activité ATPase par l'utilisation d'ATP[ $\alpha$ P32] et d'ATP[ $\gamma$ P32]

Les différents composants (indiqués par +/-) ont été incubés pendant 30mn à 55°C en présence d'ATP froid (100 $\mu$ M) supplémenté d'ATP radioactif (1 $\mu$ Ci), dans le tampon utilisé pour la réaction de synthèse de t6A (HEPES 50mM, pH8, KCl 35mM, MgCl<sub>2</sub> 5mM, MnCl<sub>2</sub> 1mM), puis 1 $\mu$ L de la réaction a été déposé sur une plaque PEI-cellulose. Les produits d'hydrolyse de l'ATP ont été séparés en chromatographie sur couche mince (CCM) dans un solvant LiCl 0,5M, acide formique 1M, sur une distance de migration de 10cm. La radioactivité a été révélée par phosphorimaging. Concentration des composants : KEOPS 1 $\mu$ M, Sua5 4 $\mu$ M, ARNt<sup>Lys(UUU)</sup>*E.coli* 12 $\mu$ M, L-thréonine 100 $\mu$ M, NaCO<sub>3</sub>=carbonate 5mM.

La position du phosphate radioactif a été représentée en orange sur la structure chimique de l'ATP pour les deux expériences (A et B) menées en parallèle dans les mêmes conditions. La position des différentes molécules est indiquée sur la gauche de chaque encart. ori= zone de dépôt. Une étoile rouge indique les molécules radioactives provenant de l'hydrolyse directe de l'ATP.

**A. Suivi des produits d'hydrolyse de l'ATP[ $\alpha$ P32]**

**B. Suivi des produits d'hydrolyse de l'ATP[ $\gamma$ P32]**

**C. Analyse des réactions B.2 à B.6 sur une distance de migration de 20cm.**

Seule la zone correspondant au dépôt, à l'ATP et au P<sub>i</sub> est montrée.



propriété. La formation d'un éventuel complexe Sua5-ARNt-KEOPS a été testée par retard sur gel. En fixant la concentration de KEOPS à une valeur intermédiaire permettant un retard des ARNt, nous avons ajouté une quantité croissante de la protéine Sua5 (**FIGURE I-2B**). La présence de la protéine ne provoque pas de retard ou d'apparition de bandes retardées supplémentaires, suggérant qu'aucune interaction stable n'advient dans ces conditions.

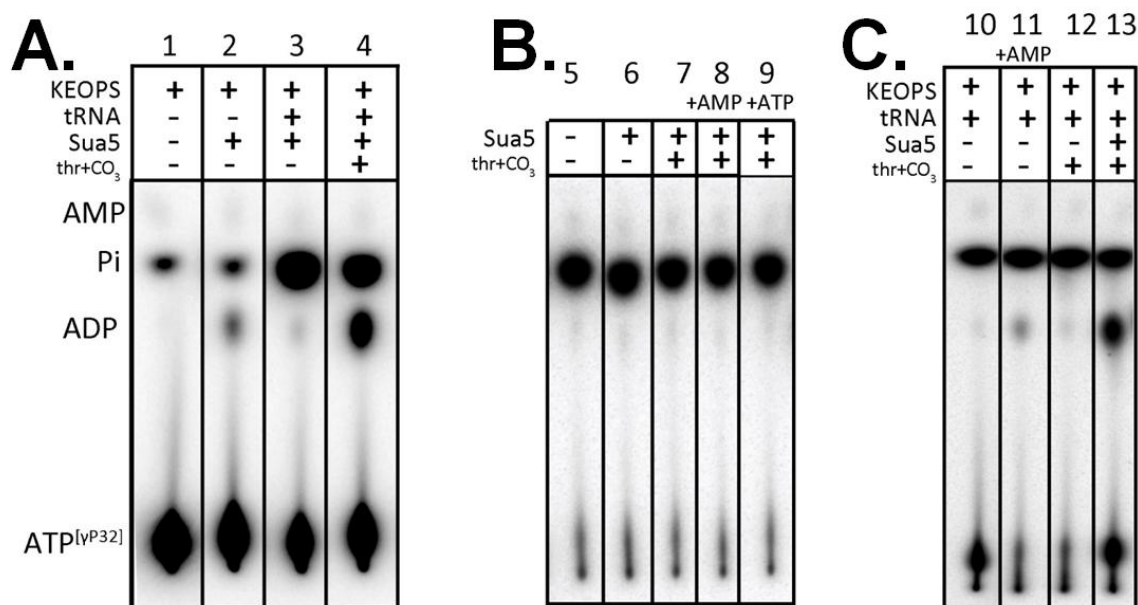
Ces différentes données suggèrent que l'interaction directe entre KEOPS et Sua5, si elle existe, doit être transitoire et indétectable par ces différentes méthodes.

### Suivi de l'activité d'hydrolyse d'ATP par Sua5 et KEOPS

Le clivage d'une molécule d'adénosine triphosphate (ATP) peut aboutir à la formation de plusieurs produits différents en fonction de la position de la liaison rompue. D'une part, un clivage entre le phosphate  $\beta$  et le phosphate  $\alpha$  conduit à la formation d'adénosine monophosphate (AMP) et d'une molécule de pyrophosphate inorganique ( $\text{PP}_i$ ) ; d'autre part, un clivage entre le phosphate  $\gamma$  et le phosphate  $\beta$  conduit à la formation d'adénosine-diphosphate (ADP) et de phosphate inorganique ( $\text{P}_i$ ). Ce dernier peut être libéré en solution ou être greffé sur un substrat, on parle alors d'activité kinase ou phosphotransférase. L'utilisation d'ATP radiomarké au niveau du phosphate  $\alpha$  ( $\text{ATP}^{[\text{P}^{32}\alpha]}$ ) nous a permis précédemment de mettre en évidence la formation d'AMP par la protéine Sua5 et d'ADP par le complexe KEOPS dans la condition de biosynthèse de  $\text{t}^6\text{A}$  (**FIGURE I-3A**). Nous avons donc également suivi ces activités en utilisant de l'ATP radiomarké au niveau du phosphate  $\gamma$  ( $\text{ATP}^{[\text{P}^{32}\gamma]}$ ), afin de suivre l'apparition éventuelle de  $\text{PP}_i$  et de  $\text{P}_i$  (**FIGURE I-3B**).

En l'absence d'ARNt, la présence du complexe KEOPS est associée à une faible production d'ADP (**dépôt A-1**), ainsi qu'à une faible production de  $\text{P}_i$  (**dépôt B-1**). La présence d'ARNt induit une forte production d'ADP (**dépôt A-2**), associée à une production équivalente de  $\text{P}_i$  libre (**dépôt B-2**). Ainsi, le  $\text{P}_i$  issu du clivage de l'ATP par KEOPS est libéré en solution.

La protéine Sua5 présente une production d'AMP fortement stimulée par la présence de carbonate et de thréonine (**dépôts A-3,4,5**). Le  $\text{PP}_i$  possédant des propriétés de migration très proches de l'ATP, ces deux molécules sont difficilement distinguables par chromatographie sur couche mince (**dépôts B-3,4,5**). Nous avons donc augmenté la distance de migration totale de la chromatographie afin d'augmenter la distance de séparation entre ces composés (**encart C**). La position du  $\text{PP}_i$  devient alors légèrement différentiable de l'ATP,



#### FIGURE I-4 : Etude de la formation *de novo* d'ADP radioactif

Dans chacune des conditions, l'ATP frais (100μM) est supplémenté d' ATP<sup>[γP32]</sup> (1μCi)

##### A. Suivi des produits radioactifs en présence des différents composants de la biosynthèse de t6A.

Le suivi des produits radioactifs a été réalisé selon les conditions de la Figure I-3B, en présence ou en absence (+/-) des différents composants (ici, [KEOPS]= 1,5μM, les autres concentrations sont identiques).

##### B. Détermination du rôle de Sua5 dans la synthèse d'ADP *de novo*

Un milieu réactionnel identique à la Figure I-3 a été additionné de 1,5μM de KEOPS, 12μM d'ARNt, et incubé pendant 30mn à 55°C. Les protéines ont été alors dénaturées par incubation à 100°C pendant 20mn. Le précipité observé a été éliminé du milieu réactionnel par centrifugation. Le surnageant (5) a été additionné de 7μM de Sua5 (6, 7, 8, 9) de thréonine et carbonate (7), d'AMP (8), ou d'ATP (9) froid à 10μM, puis incubé à nouveau 30mn à 55°C. Les produits radioactifs ont été séparés par CCM et révélés en phosphorimaging.

##### C. Détermination du rôle de KEOPS dans la synthèse d'ADP *de novo*.

Un milieu réactionnel identique à la Figure I-3 a été additionné de 1,5μM de KEOPS, 12μM d'ARNt (10, 11, 12, 13), d'AMP froid à 100μM (11), de thréonine et carbonate (12, 13), et de 4μM de Sua5 (13), puis incubé 30min à 55°C. Les produits radioactifs ont été séparés par CCM et révélés en phosphorimaging.

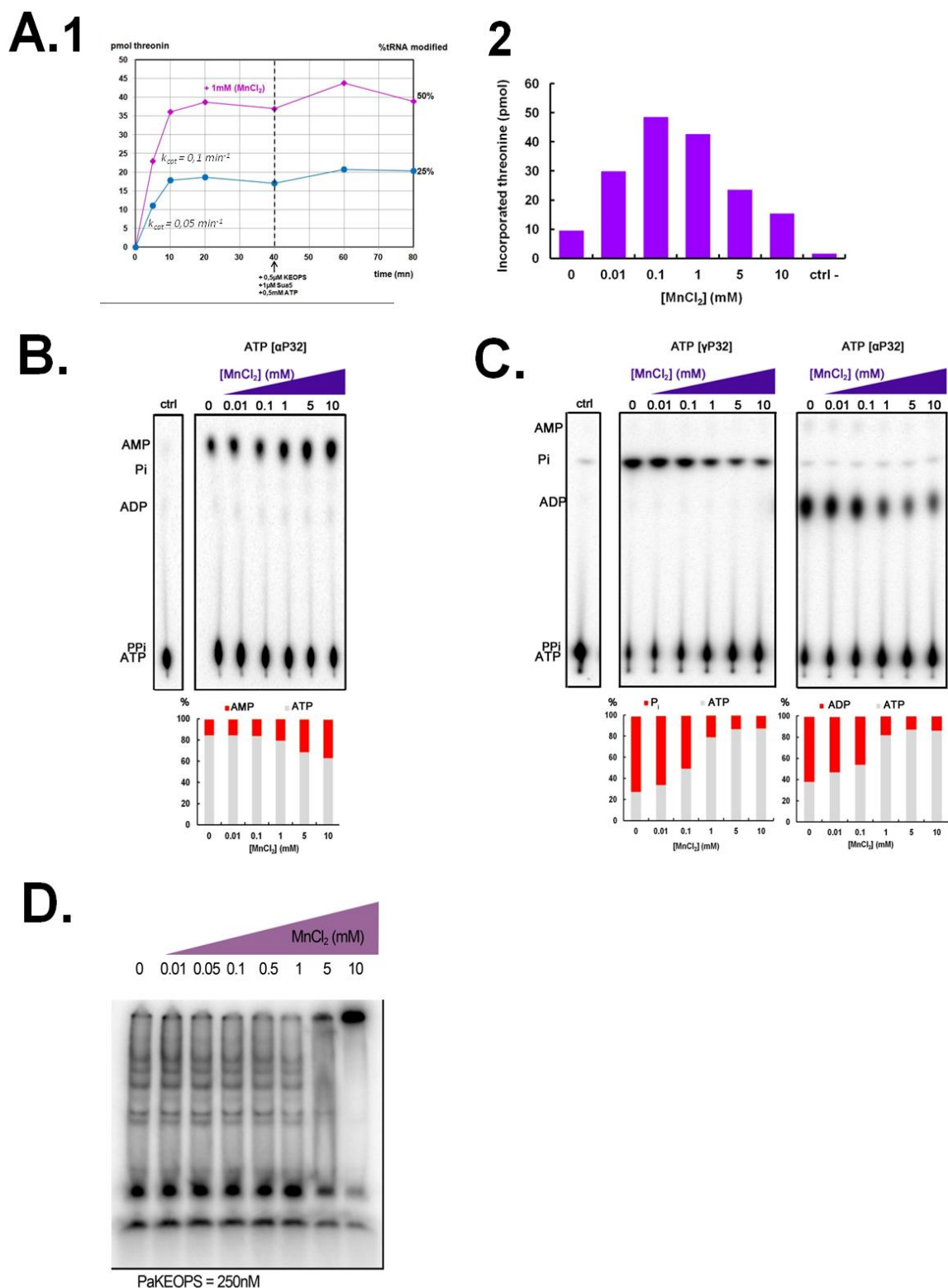
et on constate une production de PPi associée à la présence de Sua5. Ainsi, la protéine Sua5 présente une activité d'hydrolyse de l'ATP produisant de l'AMP et du PPi libéré en solution.

Lorsque tous les composants nécessaires à la biosynthèse de t<sup>6</sup>A sont réunis, on constate une consommation de la totalité de l'ATP sous la forme d'AMP, d'ADP (**dépôt A-6**) et PPi et de Pi (**dépôt B-6**), correspondant à la superposition des deux activités ATPase de KEOPS et de Sua5 (**dépôts A-2+A-5 et B-2+B-5**). De manière surprenante, dans ces conditions et en utilisant de l'ATP<sup>[P<sup>32</sup>γ]</sup>, on constate l'apparition d'un « spot » radioactif correspondant à de l'ADP (**dépôts B-6**). Or, seul le phosphate terminal de l'ATP est radiomarké dans cette expérience : la présence d'ADP radiomarké implique une formation *de novo* de cette molécule, probablement à partir d'une molécule d'AMP et d'un Pi radioactif libre.

### Origine de la synthèse *de novo* d'ADP radioactif

Afin de comprendre les modalités de formation de cet ADP, nous avons dans un premier temps incubé ensemble la protéine Sua5 et le complexe KEOPS en l'absence de leurs substrats respectifs (**FIGURE I-4A**). On constate alors la formation d'ADP radioactif en faible quantité (**dépôt 2**). Lorsque l'activité de KEOPS est stimulée par l'ajout d'ARNt mais que Sua5 n'est pas stimulée, l'intensité du spot d'ADP diminue mais reste toujours détectable (**dépôt 3**). Au bilan, il apparaît que la production de cet ADP requiert l'activation simultanée de la protéine Sua5 et du complexe KEOPS (**dépôt 4**). Nous avons alors tenté de déterminer quel composant était responsable de cette activité.

Le complexe KEOPS a été incubé en présence d'ARNt et d'ATP<sup>[P<sup>32</sup>γ]</sup> à 50°C, permettant l'hydrolyse de ce nucléotide, puis les protéines ont été dénaturées par chauffage à 100°C pendant 20mn, et éliminées par centrifugation. La totalité de la radioactivité est alors présente sous forme de Pi libre dans le surnageant (**dépôt 5**). Nous avons alors incubé ce surnageant avec la protéine Sua5 (**dépôt 6**), en présence de thréonine et de carbonate (**dépôt 7**), et en ajoutant de l'AMP non radioactif (**dépôt 8**) ou de l'ATP non radioactif (**dépôt 9**). Dans aucun des cas nous n'avons pu observer l'apparition d'ADP radioactif, suggérant que Sua5 n'est pas responsable de la production d'ADP *de novo*.



**FIGURE I-5 : Effet du MnCl<sub>2</sub> (manganèse) sur les activités de KEOPS et Sua5**

**A. Effet du manganèse sur l'activité de synthèse de t6A.**

La réaction a été réalisée selon le protocole décrit dans le Matériel et Méthodes (volume final 50μL). L'ARNt<sup>Lys(UUU)</sup> *E.coli* est à une concentration de 1,6μM (80pmol), KEOPS à 1μM, Sua5 à 2μM.

1. Suivi temporel de l'incorporation de <sup>C14</sup>L-thréonine en présence (1mM, en rose) et en l'absence (en bleu) de MnCl<sub>2</sub>. A un temps de 40min, une quantité additionnelle de KEOPS, Sua5 et ATP ont été ajouté au mélange réactionnel. Le rendement a été calculé à l'issue de la réaction, selon la formule (pmol thréonine/pmol d'ARNt)\*100. Le kcat de la réaction a été calculé en approximant la pente de la partie linéaire des courbes entre 0 et 5min (pmol thréonine/min) divisé par la quantité d'enzyme (considéré à 1μM= 50pmol)

2. Mesure de la quantité de thréonine incorporée en 30mn selon différentes concentrations de MnCl<sub>2</sub>. Les conditions sont identiques à A.1. ctrl- correspond à l'absence de KEOPS dans le milieu réactionnel. [suite de la légende p186]

Nous avons également incubé le complexe KEOPS en présence d'ARNt, d'ATP<sup>[P32γ]</sup> (dépôt 10), en ajoutant de l'AMP non radiomarké (dépôt 11). Dans ce dernier cas, on observe la présence d'un spot d'ADP radioactif. Cela suggère que le complexe KEOPS est responsable de la production d'ADP radioactif *de novo* observée précédemment. Cependant, l'intensité de ce spot n'est pas aussi intense qu'en présence de Sua5 (dépôt 13), indiquant que cette protéine pourrait produire un composé utilisé par KEOPS pour produire cet ADP radioactif avec le Pi produit dans le milieu.

### Effet du manganèse sur la machinerie de t<sup>6</sup>A :

Lors de la mise en place du système de biosynthèse *in vitro* de t<sup>6</sup>A, nous avons observé que l'addition de manganèse au milieu réactionnel (en plus du magnésium déjà présent) stimule la réaction catalysée par la protéine Sua5 et le complexe KEOPS et permet d'augmenter le rendement. Nous avons donc tenté de déterminer plus précisément l'effet de ce métal. Il est à noter que dans toutes les expériences décrites par la suite, une concentration constante de 5mM de MgCl<sub>2</sub> est présente dans le milieu réactionnel.

Nous avons dans un premier temps réalisé un suivi temporel de la production de t<sup>6</sup>A en l'absence et en présence de manganèse (Figure I-5-A1). La partie linéaire des courbes obtenues, entre 0 et 10mn, montre que la présence de 1mM de manganèse augmente la vitesse initiale de réaction d'un facteur deux. En se basant sur la concentration du complexe KEOPS (limitant par rapport à Sua5) on peut déduire très approximativement que le *turn-over* du système (Sua5+KEOPS) est de  $k_{cat}=0,1\text{min}^{-1}$  avec manganèse et  $k_{cat}=0,05\text{min}^{-1}$  en l'absence de manganèse.

Nous avons également suivi la production de t<sup>6</sup>A après 30mn (une fois le plateau atteint) à différentes concentrations de manganèse. (Figure I-5A-2). Quelle que soit la concentration testée, la production de t<sup>6</sup>A est stimulée par l'addition de cet ion, avec un optimum à 100μM de MgCl<sub>2</sub>. Au-delà de cette concentration, l'effet stimulateur diminue.

Afin de comprendre à quel niveau agit le manganèse, nous avons ensuite étudié son effet sur l'activité ATPase de Sua5 et du complexe KEOPS en suivant la production respective d'AMP et d'ADP+Pi à des concentrations croissantes de manganèse.

## FIGURE I-5 : suite de la légende p184

### A. Effet du manganèse sur l'activité de synthèse de t6A.

La réaction a été réalisée selon le protocole décrit dans le Matériel et Méthodes (volume final 50µL). L'ARNt<sup>Lys(UUU)</sup> *E.coli* est à une concentration de 1,6µM (80pmol), KEOPS à 1µM, Sua5 à 2µM.

1. Suivi temporel de l'incorporation de <sup>14</sup>C-L-thréonine en présence (1mM, en rose) et en l'absence (en bleu) de MnCl<sub>2</sub>. A un temps de 40min, une quantité additionnelle de KEOPS, Sua5 et ATP ont été ajoutée au mélange réactionnel. Le rendement a été calculé à l'issue de la réaction, selon la formule (pmol thréonine/pmol d'ARNt)\*100. Le kcat de la réaction a été calculé en approximant la pente de la partie linéaire des courbes entre 0 et 5min (pmol thréonine/min) divisé par la quantité d'enzyme (considéré à 1µM= 50pmol)

2. Mesure de la quantité de thréonine incorporée en 30mn selon différentes concentrations de MnCl<sub>2</sub>. Les conditions sont identiques à A.1. ctrl- correspond à l'absence de KEOPS dans le milieu réactionnel.

### B. Activité ATPase de Sua5 en fonction de la concentration de MnCl<sub>2</sub>

Le suivi de la formation de l'AMP a été réalisé dans les conditions de la Figure I-3. La protéine Sua5 est à une concentration de 3µM, thréonine 100µM, carbonate 5mM, ATP 100µM+1µCi ATP[aP32]. La réaction a été stoppée à 10min. ctrl= absence de Sua5. La quantité relative d'ATP et d'AMP a été quantifiée à l'aide du logiciel ImageQuant.

### C. Activité ATPase de KEOPS en fonction de la concentration en MnCl<sub>2</sub>

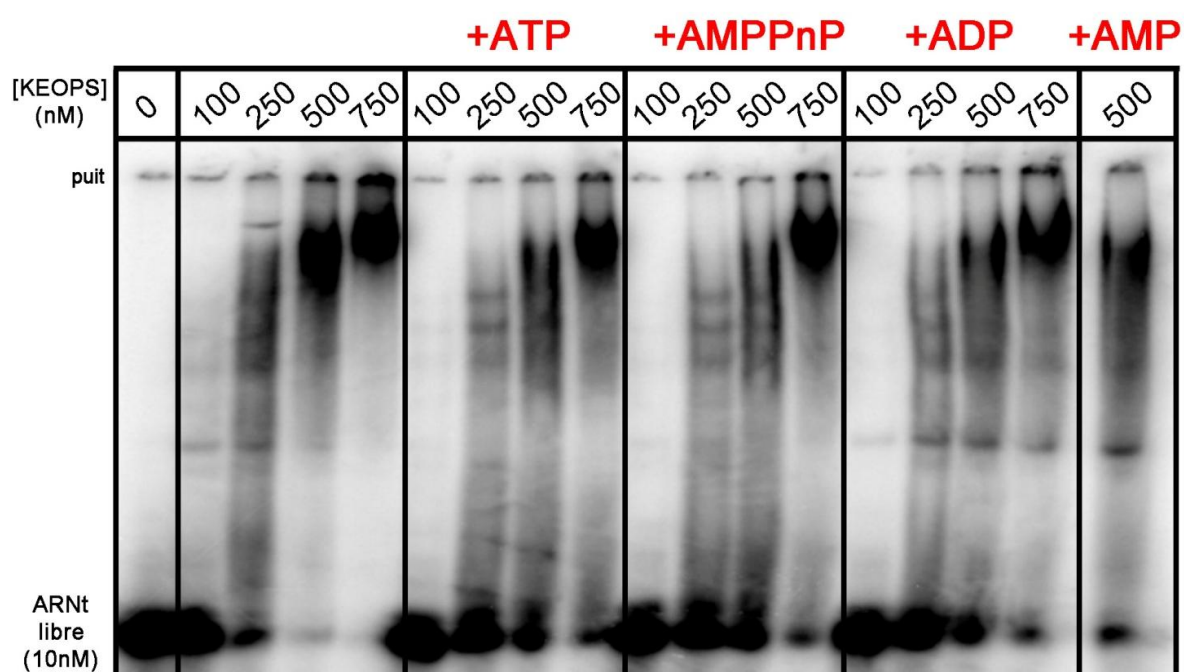
Le suivi de la formation de l'ADP + Pi a été réalisé dans les conditions de la Figure I-3. Le complexe KEOPS est à une concentration de 1µM, l'ARNt 5µM, ATP 100µM+1µCi ATP[yP32] (à gauche) ou ATP[aP32] à droite. ctrl= absence de KEOPS. La réaction a été stoppée à 10min. La quantité relative d'ATP, de Pi et d'AMP a été quantifiée à l'aide du logiciel ImageQuant.

### D. Fixation de l'ARNt par KEOPS en retard sur gel en fonction de la concentration en MnCl<sub>2</sub>

Le protocole de retard sur gel est identique à celui utilisé dans la Figure I-2. L'ARNt utilisé comme sonde est un ARNt<sup>Lys(UUU)</sup> transcrit *in vitro*, à une concentration de 10nM. Le complexe KEOPS est à une concentration fixe de 250nM. La concentration en MnCl<sub>2</sub> dans chaque réaction est indiquée en haut du gel en mM.

Dans le cas de la protéine Sua5 (**FIGURE I-5B**), aucun effet n'est détectable à 10 $\mu$ M ou 100 $\mu$ M de manganèse. En revanche, une concentration égale ou supérieure à 1mM stimule jusqu'à 4 fois la production d'AMP par Sua5. Cependant, cet optimum ne correspond pas à la concentration optimale observée pour la réaction t6A (100 $\mu$ M). Dans le cas du complexe KEOPS (**FIGURE I-5C**) on constate au contraire une inhibition de l'activité ATPase en présence de manganèse. A des concentrations de 5 et 10 mM, l'activité est diminuée d'un facteur 4.

Enfin, le manganèse étant un ion divalent pouvant intervenir dans l'interaction intermoléculaire entre acide nucléique et protéine, nous avons également étudié l'effet du manganèse sur la fixation des ARNt par le complexe KEOPS en retard sur gel (**FIGURE I-5D**). Une concentration fixe de KEOPS et d'ARNt radiomarqués a été incubée sur glace, pendant 15mn, en présence de différentes concentrations de manganèse, puis les complexes nucléoprotéiques ont été séparés sur gel de polyacrylamide en condition native. De 0 à 1mM, l'addition de manganèse n'a pas d'effet détectable. En revanche, à 5mM et 10mM de manganèse, c'est-à-dire aux concentrations où l'activité ATPase du complexe KEOPS est la plus fortement inhibée, on constate une forte perturbation de migration correspondant à une agrégation de complexes nucléoprotéiques au niveau des puits du gel. Ainsi, cela suggère que l'inhibition de l'activité d'hydrolyse de l'ATP causée par le manganèse pourrait être causée par une perturbation structurale au niveau de l'interaction entre le complexe KEOPS et l'ARNt.



**FIGURE I-6 : Fixation de l'ARNt par KEOPS en présence de différents nucléotides**

Le protocole de retard sur gel est identique à celui utilisé dans la Figure I-5-D.

La concentration du complexe KEOPS est indiquée en haut du gel pour chaque réaction, en nM. Une quantité de 1mM de chacun des nucléotides a été ajoutée, et le mélange a été incubé à 55°C pendant 15mn avant séparation des complexes nucléoprotéiques par électrophorèse.



### Rôle de l'hydrolyse de l'ATP sur la fixation aux ARNt:

Sachant donc que l'activité d'hydrolyse de l'ATP par le complexe KEOPS requiert la présence d'ARNt dans le milieu réactionnel, et probablement d'une fixation directe comme le montre l'observation précédente, nous avons alors cherché à savoir si, en retour, l'hydrolyse de l'ATP modifie les capacités de fixation aux ARNt du complexe KEOPS.

Nous avons incubé différentes concentrations de complexe KEOPS avec un ARNt radiomarké, en présence respectivement de 1mM d'ATP, d'AMPPNP (un analogue non hydrolysable de l'ATP), d'ADP et d'AMP ou en l'absence de nucléotide, dans les conditions utilisées pour le suivi de l'activité ATPase (incubation à 55°C). Au bout de 15mn, les réactions ont été stoppées sur glace et les complexes nucléo-protéiques ont été séparés sur gel de polyacrylamide en condition native (**FIGURE I-6**)

On constate que la présence d'ATP semble diminuer légèrement la fixation du complexe KEOPS aux ARNt. Le résultat est identique avec de l'AMPPNP et l'ADP, suggérant que l'effet observé provient de la fixation du nucléotide et non de son hydrolyse.

Ainsi, alors que l'hydrolyse de l'ATP dépend strictement de la présence d'ARNt, nos résultats ne permettent pas conclure si l'hydrolyse de l'ATP affecte en retour la fixation aux ARNt.



## Modalités de l'hydrolyse de l'ATP par Sua5 et KEOPS : étude enzymatique préliminaire

La mise en évidence des activités ATPases précédemment décrites permettent de mieux comprendre le fonctionnement de la machinerie de synthèse de t<sup>6</sup>A chez *P.abysyi*. Afin d'en déterminer les modalités, nous avons alors fait varier certains paramètres réactionnels. Dans chacune des conditions décrites par la suite, l'ATP non radiomarké présent à une concentration de 100μM est supplémenté de 1μCi d'ATP<sup>[P32α]</sup>.

### A - Modalités de la production d'AMP et par la protéine Sua5...

... en fonction du temps (FIGURE I-7C) :

Dans un premier temps, nous avons suivi la production d'AMP par Sua5 en fonction du temps, en présence de thréonine et de carbonate, à 50°C. Entre 0 et 15mn, la courbe obtenue présente une progression linéaire, puis atteint une phase stationnaire à partir d'environ 20mn, où la production d'AMP se stabilise.

En considérant que la concentration des substrats correspond à un large excès (5mM de carbonate et 50μM de thréonine pour 0,5μM d'enzyme), on suppose que la vitesse au niveau de la partie linéaire, ou vitesse initiale, correspond à la vitesse maximale de la réaction d'hydrolyse de l'ATP dans ces conditions ( $V_{max}$ ). Nous en déduisons alors que  $V_{max} = 3,2\mu\text{M}/\text{minutes}$  (soit 80fmol/min). D'après cette valeur, le « turn-over » ( $k_{cat}$ ) de Sua5 vis-à-vis de la consommation d'ATP est estimé à  $k_{cat(Sua5)} = 6,4 \text{ min}^{-1}$ , autrement dit une molécule de Sua5 hydrolyse 6,4 molécules d'ATP en 1 minute.

Lorsque le plateau est atteint, seul 40% de l'ATP a été consommé. Nous avons donc également suivi la production d'AMP en faisant varier respectivement la concentration de l'enzyme, Sua5, et la concentration du substrat, la L-thréonine, afin de déterminer les concentrations nécessaires pour l'hydrolyse totale de l'ATP. Ces deux suivis, où la réaction a été stoppée après 30mn d'incubation (correspondant au début de la phase stationnaire) fournissent un aperçu préliminaire du comportement de Sua5 dans ces conditions in vitro mais ne permettent pas de déterminer les constantes enzymatiques de la réaction ( $V_{max}$  et  $K_m$ ).



*... en fonction de la concentration de Sua5 (FIGURE I-7A)*

A concentration fixe de thréonine (50 $\mu$ M, comme précédemment), l'activité de Sua5 est détectable dès 100nM de Sua5 et augmente progressivement en fonction de la quantité de protéines. Au-delà de 2,5 $\mu$ M de protéines, la totalité des 100 $\mu$ M d'ATP sont consommés en 30mn d'incubation.

*... en fonction de la concentration de thréonine (FIGURE I-7D)*

A concentration fixe de Sua5 (1 $\mu$ M), la production d'AMP par Sua5 est détectable dès 500nM de thréonine, et une concentration d'au moins 100 $\mu$ M de thréonine est nécessaire pour obtenir une consommation totale des 100 $\mu$ M d'ATP. Ainsi, un large excès de thréonine par rapport à l'enzyme est nécessaire pour obtenir une activation totale, suggérant que l'affinité pour ce substrat est de l'ordre du micromolaire.

*... en fonction de la température (FIGURE I-7B)*

Enfin, nous avons voulu connaître l'influence de la température sur cette activité. Nous avons donc suivi la production d'AMP après 30mn d'incubation à différentes températures d'incubation (Figure C1 – B). Il apparaît que la température optimale se situe **entre 57°C et 67°C** dans ces conditions, légèrement supérieure à la température optimale déterminée pour l'activité de biosynthèse de t<sup>6</sup>A (55°C)

**B - Modalités de la production d'ADP et par le complexe KEOPS...**

Nous avons effectué un suivi similaire de l'activité d'hydrolyse d'ATP en ADP par le complexe KEOPS, en présence d'ARNt<sup>Lys(UUU)</sup> d'*E.coli* produit in vivo, qui est un substrat de la réaction de biosynthèse de t<sup>6</sup>A.

*...en fonction du temps (FIGURE I-8B)*

La courbe de production d'ADP en fonction du temps, à une concentration fixe de KEOPS (100nM) et d'ARNt (1 $\mu$ M), présente une phase linéaire qui se situe entre 0 et 50mn, suivie d'une phase stationnaire. A partir de la pente de la phase linéaire, il est possible de donner une première estimation très approximative du  $k_{cat}$ , environ 8 à 10min<sup>-1</sup>.

*...en fonction de la concentration de KEOPS (FIGURE I-8A)*

Nous avons par la suite fait varier la concentration du complexe KEOPS, en stoppant la réaction à 30mn d'incubation, à concentration fixe d'ARNt (1 $\mu$ M). Il est à noter que cette expérience a été réalisée avec 10 $\mu$ M d'ATP final dans la réaction au lieu des 100 $\mu$ M utilisés dans les autres expériences.

Les différentes conditions sont indiquées dans chaque encart, et dans le texte.



On constate l'apparition d'ADP à de très faibles concentrations de KEOPS (25nM). Aux valeurs les plus basses de la concentration d'enzyme (lorsque l'ATP est encore en très large excès), la quantité d'ADP formée est proportionnelle à la concentration de KEOPS. Le  $k_{cat}$  approximatif obtenu d'après ces valeurs est d'environ  $5 \text{ min}^{-1}$ , comparable à celui estimé précédemment (par exemple,  $3 \mu\text{M}$  d'ATP sont produits par 25nM de KEOPS en 30mn, ce qui donne 4 molécules d'ADP par molécule de KEOPS par minute).

A partir de  $1 \mu\text{M}$  de KEOPS, la totalité de l'ATP est consommée en ADP. De manière surprenante, au-delà de cette concentration, l'ADP est lui-même converti en un produit formant un « spot » radioactif migrant à la position théorique du phosphate inorganique libre ( $\text{Pi}$ ). Ce spot devient même majoritaire à partir de  $10 \mu\text{M}$  de KEOPS. Dans cette réaction, le phosphore  $\text{P}^{32}$  radioactif de l'ATP est placé en position alpha, attaché au ribose. La production d'un  $\text{Pi}$  radioactif sous forme libre à partir d'ADP implique donc deux clivages, l'un entre le phosphate  $\beta$  et  $\alpha$ , l'autre entre le phosphate  $\alpha$  et le ribose. Le spot radioactif observé pourrait également correspondre à un dérivé nucléotidique possédant des propriétés de migration similaire au phosphate libre.

### *...en fonction de la concentration d'ARNt (FIGURE I-8C)*

Enfin, nous avons mesuré l'influence de la concentration en ARNt sur l'activité de KEOPS.

Dans un premier essai, après 30mn d'incubation et pour une concentration fixe de KEOPS (100nM), nous avons observé que la production d'ADP est maximale dès 50nM d'ARNt, traduisant une très forte affinité.

Nous avons donc réalisé la même expérience stoppant la réaction à 10mn d'incubation, afin de se placer dans les conditions de vitesse initiale et déterminer les constantes enzymatiques du complexe KEOPS pour le substrat ARNt. En utilisant la représentation en double inverse de Lineweaver-Burke, la vitesse maximale d'hydrolyse déduite est  $V_{max} = 17,3 \text{ fmol/min}$  et le  $K_m$  pour les ARNt = 11nM.

Le turn over  $k_{cat}$  calculé est donc de  $6 \text{ min}^{-1}$ , en accord avec les estimations précédentes, et comparable à celui estimé pour Sua5.





## Discussion complémentaire

### Le domaine N-terminal de Sua5, élément régulateur du domaine YrdC

Nous avons voulu comprendre le rôle du domaine C-terminale « SUA5 » retrouvé fusionné à YrdC chez certaines espèces de Bactéries, d'Archées et d'Eucaryotes, notamment chez *P.abyssii* et *S.cerevisiae*. Le domaine catalytique YrdC de la protéine de Sua5 de *P.abyssii* exprimé sans ce domaine ne présente plus aucune activité, suggérant que sa présence est requise pour la fonction de cette protéine. De manière intéressante, le domaine YrdC est retrouvé majoritairement dans la fraction insoluble lors de son expression dans *E.coli*. Le domaine SUA5 pourrait donc avoir un rôle structural et stabilisateur, permettant à la protéine entière et/ou à son site actif d'adopter une conformation correcte. Il a été suggéré qu'il puisse également jouer un rôle dans la cinétique d'hydrolyse de l'ATP, et pourrait éventuellement fixer voire hydrolyser des nucléotides (LAUHON ET AL, 2012). Nous n'avons observé aucune activité d'hydrolyse dans nos tests ATPase, mais il serait intéressant d'étudier les propriétés de fixation aux nucléotides de ce domaine.

L'absence d'interaction fonctionnelle et physique que nous avons observée entre nos deux protéines tronquées suggère que le « linker » de 20 résidus constitue un élément fondamental pour l'interaction entre les deux domaines et peut-être dans l'activité de la protéine. Ce linker n'apparaît pas dans la structure décrite de la protéine Sua5 de *Sulfolobus tokodaii* (AGARI ET AL. 2008), en raison de sa flexibilité. Récemment, nos collaborateurs (W. Zhang, H.V.Tilbergh) ont résolu la structure de Sua5 de *Pyrococcus abyssi* (qui présente dans les conditions utilisées une molécule de L-thréonine et de PPi dans le site actif du domaine YrdC), où apparaît la position exacte de ce linker (DONNEES NON PUBLIEES). De manière intéressante, le linker forme une boucle qui se projette au cœur du site actif du domaine YrdC, au niveau de la position attendue de l'AMP (par comparaison avec la structure de Sua5 de *S.tokodaii*). Ainsi, il pourrait former une « boucle de régulation » (*gating loop*) contrôlant la formation de l'intermédiaire. Il serait alors intéressant d'effectuer des mutations ponctuelles ou des délétions moins importantes au niveau du domaine C-ter et de ce linker,

afin de déterminer dans quelle mesure il intervient dans l'activité de Sua5, ce qui pourrait fournir des indications sur le mécanisme catalytique de formation de l'intermédiaire TC-AMP.

Cependant, la présence d'une telle boucle (et du domaine C-ter associé) n'apparaît pas strictement nécessaire à l'activité du domaine YrdC d'une manière générale, puisque de nombreuses espèces de Bactéries et d'Archées ne le possèdent pas. La présence d'un tel domaine semble donc refléter l'acquisition d'un mécanisme de contrôle supplémentaire apparu au cours de l'évolution chez certaines espèces. Sa répartition au sein du monde vivant est très intrigante, et pourrait témoigner de transferts de gènes horizontaux entre Bactéries, Archées et Eucaryotes. Il serait intéressant d'étudier la phylogénie exhaustive du domaine SUA5 et du domaine YrdC de manière indépendante, afin de retracer l'histoire de cette famille et de ces domaines protéiques.

Ce domaine additionnel pourrait être également impliqué dans une autre fonction que la synthèse de t<sup>6</sup>A. La recherche de protéines présentant une distribution identique pourrait ouvrir des voies pour trouver d'éventuels partenaires.

**Enfin, la présence spécifique du domaine SUA5 chez les champignons et chez de très nombreux Eucaryotes unicellulaires parasites de l'Homme qui sont à l'origine de graves maladies (paludisme, maladie du sommeil, leishmaniose) pourrait en faire une cible de choix pour développer de nouvelles molécules anti-parasitaires et anti-fongiques. Il est donc essentiel de comprendre le rôle et le fonctionnement de ce domaine C-terminal au niveau biochimique.**

### Sua5 et KEOPS, des composants indépendants ?

Nos données complémentaires suggèrent que la protéine Sua5 et le complexe KEOPS ne forment pas d'interaction stable *in vitro*. De plus, nos résultats de complémentation *in vitro* suggèrent que le TC-AMP produit par Sua5 est directement libéré dans le milieu réactionnel. Cependant, il semble étonnant que les deux étapes de cette réaction soient totalement découplées, au vu de l'instabilité d'un tel intermédiaire (entre 3,5 et 16mn selon

les conditions à 37°C, LAUHON ET AL., 2012), en particulier aux hautes températures dans le cas de *Pyrococcus abyssi*. Ainsi, nous ne pouvons pas exclure qu'une interaction transitoire puisse effectivement avoir lieu afin de permettre à Sua5 de transmettre efficacement l'intermédiaire au complexe KEOPS. Si une telle interaction existe, il est probable qu'elle ait lieu entre Sua5 et la protéine Kae1. En effet, ce sont les deux seuls composants universels du système, et de plus, chez les carbamoyl-transférases TobZ et HypF, les deux domaines YrdC et Kae1, qui sont fusionnés, forment un tunnel par lequel transite l'intermédiaire réactionnel (PARTHIER ET AL. 2012). Ainsi, des études interactionnelles plus sensibles, par la technique de résonance plasmonique de surface par exemple, pourraient permettre de mettre en évidence une telle interaction.

### Origine et rôle de l'activité ATPase de KEOPS

Dans les conditions de biosynthèse de t<sup>6</sup>A, nous avons montré que le complexe KEOPS présente une activité d'hydrolyse de l'ATP, produisant de l'ADP et du Pi libre en présence d'ARNt. A forte concentration du complexe KEOPS, lorsque l'ATP n'est plus disponible, nous avons également remarqué un produit radioactif supplémentaire provenant de la consommation de l'ADP par le complexe KEOPS. Ce produit a les mêmes propriétés de migration que le Pi libre. La détermination de la nature exacte de ce composé pourrait fournir des informations sur les modalités de clivage de l'ATP par le complexe KEOPS. Cependant, on peut penser que cette activité n'a pas de réalité physiologique, car elle apparaît lorsque l'ATP est limitant et que la concentration de KEOPS est très importante.

Deux protéines susceptibles d'être responsable de l'activité ATPase de KEOPS : d'une part, la protéine Kae1, dont la structure montre qu'un nucléotide peut se placer au niveau de son site actif (HECKER ET AL., 2007), d'autre part la protéine Bud32 qui possède en effet une P-loop permettant la fixation de l'ATP (FACCHIN ET AL., 2002, HECKER ET AL., 2008). L'étude des différents composants du complexe KEOPS, que nous aborderons dans la seconde partie de ce manuscrit, permettra de déterminer quelle sous-unité est responsable de cette activité.

De plus, l'activité d'hydrolyse de l'ATP en ADP+Pi est directement liée à la présence d'ARNt dans le milieu réactionnel. Elle pourrait donc jouer un rôle dans la fixation voire dans la reconnaissance du substrat. Nos résultats en retard sur gel en présence de différents

nucléotides ne permettent pas de conclure quant à cette hypothèse, et il serait intéressant d'utiliser des techniques plus sensibles telle que la microcalorimétrie afin d'étudier les interactions ARNt/KEOPS en fonction de la présence de nucléotide. Nous reviendrons également sur le rôle possible de cette activité ATPase dans la seconde partie du manuscrit.

### KEOPS produit de l'ADP à partir d'AMP et de Pi

Nous avons mise en évidence une activité de production *de novo* d'ADP par la machinerie de synthèse de t<sup>6</sup>A, qui n'entre pas dans le mécanisme réactionnel que nous avons proposé. Nous avons montré que KEOPS semble à l'origine de cette activité, car il peut produire une molécule d'ADP en présence d'AMP et de Pi. Cependant, la présence simultanée de Sua5, de carbonate et de thréonine, les précurseurs de t<sup>6</sup>A, semblent favoriser cette production d'ADP. Ainsi, cette activité pourrait révéler de façon indirecte une partie du mécanisme catalytique et la présence de l'intermédiaire TC-AMP produit par la protéine Sua5. Le complexe KEOPS fixerait le TC-AMP, grefferait le groupement TC sur l'ARNt, et la molécule d'AMP restant dans le site actif serait alors combinée à un phosphate inorganique libre (et radioactif dans nos expériences), produisant une molécule d'ADP qui serait relarguée. Il serait intéressant d'isoler l'intermédiaire TC-AMP afin de démontrer qu'il est bien à l'origine de l'ADP formé par KEOPS.

### Le manganèse, un cofacteur impliqué à plusieurs niveaux ?

Un dernier élément intéressant que nous avons identifié lors de cette étude est l'effet de l'ion divalent manganèse : sa présence (100μM) stimule la réaction de synthèse de t<sup>6</sup>A. L'activité ATPase de Sua5, qui reflète la synthèse de l'intermédiaire TC-AMP, semble également stimulée dans une moindre mesure par la présence du manganèse. Or, il a été montré qu'un atome de manganèse est présent dans le site actif du domaine YrdC de la protéine TobZ (PARTHIER ET AL, 2012). Ainsi, le manganèse pourrait jouer un rôle dans la production ou la stabilisation de l'intermédiaire réactionnel par Sua5. De même, il est possible qu'un ion manganèse se place au niveau du site de transfert du groupement TC sur l'ARNt dans l'étape terminale de la réaction, améliorant la catalyse.

Dans les mêmes gammes de concentration de manganèse, l'activité ATPase de KEOPS semble au contraire inhibée. Or la protéine Bud32 a été décrite comme une kinase stimulée par la présence du manganèse (HECKER ET AL., 2008). Il semble donc surprenant que le manganèse ait un effet inhibiteur sur KEOPS. La compréhension du rôle de Bud32 dans le complexe KEOPS permettrait donc de savoir à quel niveau intervient cet ion divalent.

Enfin, aux fortes concentrations de manganèse (de l'ordre du millimolaire), on constate que la perte totale d'activité ATPase de KEOPS est corrélée à une forte perturbation de la fixation aux ARNt. Ainsi, on peut penser que le manganèse joue un rôle dans la structuration de l'ARNt, ou dans les interactions entre ARNt et KEOPS, et donc dans la fixation du substrat.

Le manganèse pourrait donc intervenir à différents niveaux dans le système complexe de biosynthèse de t<sup>6</sup>A, nos données préliminaires ne permettent pas d'expliquer précisément son rôle et l'effet stimulateur qu'il semble avoir sur la réaction globale, mais ouvrent différentes pistes à explorer.

## La détermination des paramètres enzymatiques: vers la compréhension du mécanisme réactionnel

Récemment, certains paramètres enzymatiques du système de biosynthèse de t<sup>6</sup>A bactérien ont été déterminés à partir de protéines purifiées de *B.subtilis* (LAUHON ET AL., 2012). La détermination des paramètres enzymatiques du système que nous avons mis en place permettrait de mettre en évidence certaines similitudes ou différences entre les machineries Bactériennes et Archéennes, et de mieux comprendre le mécanisme réactionnel en jeu.

Nous avons débuté cette caractérisation pour certaines de ces activités. Ainsi, nous avons montré que le  $k_{cat}$  de la biosynthèse de t<sup>6</sup>A est d'environ 0,1min<sup>-1</sup>, proche de celle déterminée pour le système Bactérien qui est de 0,4min<sup>-1</sup> (LAUHON ET AL., 2012). Ainsi, ces deux systèmes enzymatiques pourtant très différents semblent fonctionner avec une efficacité similaire. Ces valeurs obtenues peuvent sembler faible, et s'expliquer par des conditions *in vitro* qui ne reflètent pas les conditions optimales *in vivo*, et d'autre part, les ARNt étant des molécules stables dans le temps (un turn-over estimé à plusieurs dizaines d'heures, PHIZICKY ET AL., 2010), il n'est pas nécessaire d'avoir un turn over très rapide pour

leur enzyme de modification. En effet, d'autres enzymes de modification ont un turn over du même ordre de grandeur ( $0,5\text{min}^{-1}$  pour Trm5 de *M.jannaschii*, qui catalyse la formation de  $\text{m}^1\text{G}_{37}$ , CHRISTIAN ET AL. 2006)

Les turn-over ( $k_{\text{cat}}$ ) déterminés pour l'hydrolyse de l'ATP par Sua5 d'une part et KEOPS d'autre part sont également comparables et proche du  $k_{\text{cat}}$  obtenu pour la synthèse de t6A. Bien que nous n'ayions pas déterminé le  $K_m$  (qui traduit l'affinité pour le substrat) de Sua5 pour la thréonine, il semble nettement plus élevé (de l'ordre du  $\mu\text{M}$ ) que celui de KEOPS pour les ARNt ( $K_m=11\text{nM}$ ). Cela permet de suggérer que l'efficacité enzymatique de KEOPS (le rapport  $k_{\text{cat}} / K_m$ ) pour l'hydrolyse de l'ATP est plus grande que celle de Sua5, et que cette dernière doit constituer l'élément limitant du système de biosynthèse de t6A.

Différentes perspectives s'ouvrent dans la caractérisation enzymatique plus poussée de ce système :

- L'hydrolyse de l'ATP étant liée à la présence d'ARNt, il serait intéressant de déterminer les constantes enzymatiques pour différentes espèces d'ARNt, substrats et non substrats de la réaction, ce qui pourrait révéler un mécanisme de sélection du substrat par le complexe KEOPS.
- La protéine Sua5 produit l'intermédiaire TC-AMP à partir de thréonine, de carbonate et d'ATP. Il a été proposé que cette formation passe par une première étape de condensation de la thréonine et du carbonate en carboxy-thréonine de manière ATP indépendante (LAUHON ET AL, 2012.), suivi de l'adénylation de ce composé par l'hydrolyse d'une molécule d'ATP. La détermination précise du nombre d'ATP consommé par molécule de TC-AMP produit par suivi enzymatique permettrait de connaître les modalités de cette réaction.
- Enfin, la comparaison des paramètres enzymatiques entre les systèmes Bactériens, Archéens et Eucaryotes permettrait de révéler des points communs mais aussi les divergences qui peuvent exister entre ces systèmes.

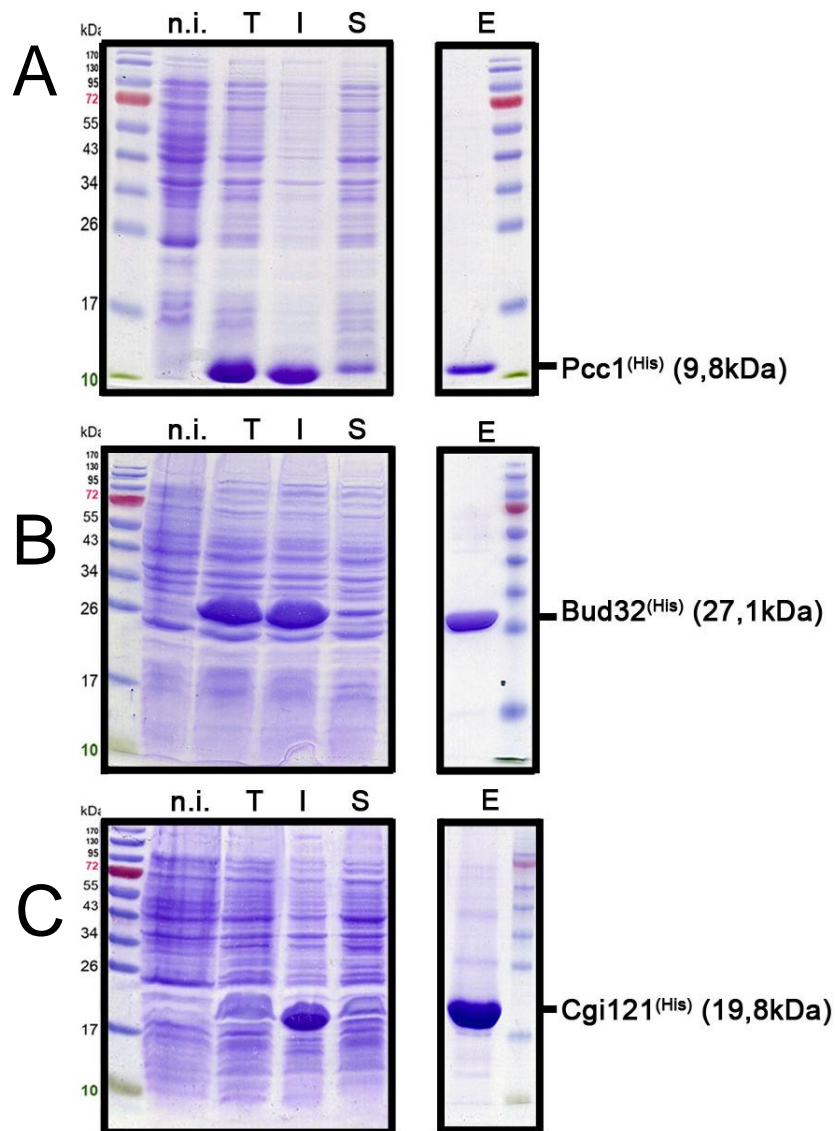
## Conclusion :

Les résultats principaux et complémentaires que nous avons obtenus constituent ainsi des pistes vers la compréhension de la voie de biosynthèse de t<sup>6</sup>A de *Pyrococcus abyssi*. Nous avons pu mettre en évidence que cette biosynthèse fait appel à une machinerie complexe, comprenant 5 protéines (Sua5, Kae1, Bud32, Pcc1 Cgi121), qui utilisent de nombreux substrats et cofacteurs (ATP, thréonine, carbonate, ARNt, manganèse) et où nombreuses activités enzymatiques sont présentes simultanément (hydrolyse de l'ATP, fixation à l'ARNt). La comparaison de nos résultats et des données disponibles dans la littérature nous a permis d'élaborer un modèle globale pour cette réaction. Alors que la protéine Sua5 et ses orthologues fonctionnent de manière isolée pour produire l'intermédiaire réactionnel à partir de la thréonine et du carbonate, il apparait que la protéine Kae1 nécessite la présence de partenaires spécifiques pour effectuer la seconde étape de la réaction, que ce soit chez les Bactéries, les Archées et les Eucaryotes. Il est donc nécessaire de comprendre l'importance de chacun de ces partenaires afin de pouvoir comparer les différents systèmes. Cette étude fera l'objet de la seconde partie de ce manuscrit, où nous essaierons de répondre à la question suivante: quelle est le rôle des différentes sous-unités du complexe KEOPS dans la biosynthèse de t<sup>6</sup>A ?





## **PARTIE II - Etude du rôle des sous-unités de KEOPS**



**FIGURE II-1 : Expression et purification des sous unités Pcc1, Cgi121 et Bud32 de *Pyrococcus abyssi***

Les protéines recombinantes portant une étiquette hexa-histidine en position C-terminale (-His) ont été sur-exprimée dans une souche Rosetta2(DE3)pLysS d'*Escherichia coli* par induction à l'IPTG. Après induction, les cellules ont été resuspendues dans un tampon de purification et lysées par sonication. La fraction soluble du lysat total a été séparée de la fraction insoluble par centrifugation, puis a été purifiée par chromatographie d'affinité. Un alicot de chaque fraction a été analysé par SDS PAGE. La masse moléculaire des marqueurs protéiques est indiquée sur la gauche de la figure, en kiloDalton (kDa). La masse moléculaire théorique de chaque protéine recombinante est indiquée sur la droite de la figure entre parenthèse. n.i. = lysat cellulaire totale non induit à l'IPTG. T= lysat cellulaire après induction à l'IPTG. I= fraction insoluble du lysat. S= fraction soluble du lysat. E= fraction d'élution après chromatographie d'affinité.

- A. Purification de la protéine Pcc1-His (PAB3073-His)
- B. Purification de la protéine Bud32-His (PAB1047-His)
- C. Purification de la protéine Cgi121-His (PAB1522-His)

## Résultats

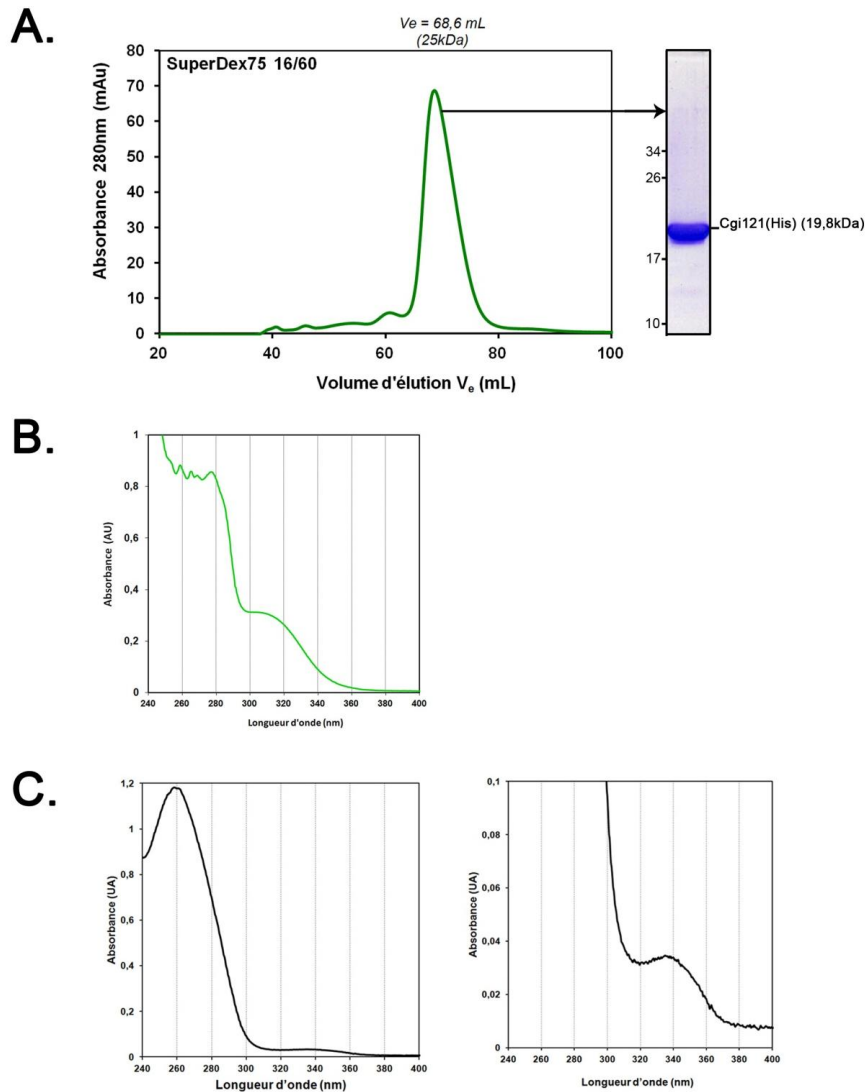
Précédemment, nous avons montré que le complexe KEOPS de *Pyrococcus abyssi* présente trois activités identifiées *in vitro*: biosynthèse de t<sup>6</sup>A (en présence de Sua5), fixation à l'ARNt, et hydrolyse de l'ATP en ADP+P<sub>i</sub> stimulée par la présence d'ARNt. Afin de déterminer l'importance et le rôle de chacun des composants du complexe KEOPS dans ces trois activités, nous avons dans un premier temps cherché à obtenir de manière isolée les quatre sous-unités Pcc1, Kae1, Bud32, Cgi121 de *Pyrococcus abyssi*.

### Purification des protéines isolées Pcc1, Kae1, Bud32, Cgi121

Les gènes codant respectivement ces quatre protéines ont été clonés séparément dans un vecteur d'expression, fusionné avec une séquence codant une étiquette hexa-histidine en position 3' (se retrouvant une fois exprimée en position C-terminale de la protéine, et notée <sup>-His</sup>) et sur-exprimés par la souche Rosetta2(DE3)pLysS d'*Escherichia coli*, selon le protocole utilisé pour la purification du complexe KEOPS (PERROCHIA ET AL. 2013, Matériel et Méthodes)

La purification et la caractérisation de Kae1<sup>-His</sup> a fait l'objet d'une précédente publication (HECKER ET AL, 2007), dont les résultats sont brièvement rappelés. Kae1<sup>-His</sup> s'exprime de manière soluble, et se trouve sous forme monomérique *in vitro*. Cette protéine contient un atome de fer au niveau de son site catalytique, provoquant une coloration rose associée à un maximum d'absorbance caractéristique à 500nm, que l'on retrouve également lors de la purification du complexe KEOPS entier. Contrairement au complexe entier, aucun ARN n'a été détecté associé à la protéine lors de sa purification.

Les protéines Pcc1<sup>-His</sup> (FIGURE II-1A), Bud32<sup>-His</sup> (FIGURE II-1B), et Cgi121<sup>-His</sup> (FIGURE II-1C), sur-exprimées dans *E.coli* sont retrouvées majoritairement dans la fraction insoluble du lysat cellulaire (I), mais également en quantité détectable dans la fraction soluble (S). Chaque protéine recombinante a alors été purifiée (E) par une première étape de chromatographie d'affinité sur résine d'agarose couplée avec du nickel (Ni-NTA).



## FIGURE II-2 : Purification de la protéine Cgi121<sup>-His</sup> par filtration sur gel

A l'issue de l'étape de chromatographie d'affinité (Figure II-1-C), la protéine recombinante Cgi121-His a été soumise à une seconde étape de purification de filtration sur gel sur une colonne SuperDex75 16/60. Après analyse SDS-PAGE, les fractions d'élution contenant la protéine purifiée ont été réunies, concentrées, et le spectre d'absorption de la solution protéique a été mesuré entre 200nm et 1000nm.

### A. Chromatogramme obtenu après filtration sur gel et analyse SDS-PAGE du pic d'élution majoritaire.

L'axe des abscisses correspond au volume d'élution des fractions en sortie de colonne, et l'axe des ordonnées à l'absorbance à 280nm des fractions correspondantes en milli-unité d'absorbance. Le volume d'élution du pic majoritaire à 280nm est indiqué en haut de la figure, et la masse moléculaire correspondante calculée à l'aide de marqueurs protéiques globulaires est indiquée entre parenthèse. L'analyse SDS-PAGE de la fraction d'élution correspondante à ce pic est montrée à droite du chromatogramme. Les masses moléculaires des marqueurs protéiques est indiquée en kDa sur la gauche de l'encart.

### B. Spectre d'absorption de de Cgi121-His purifiée.

L'axe des abscisses correspond à la longueur d'onde en nm, seul l'intervalle entre 240 et 400nm est représenté. L'axe des ordonnées correspond à l'absorbance, en unité d'absorbance (UA)

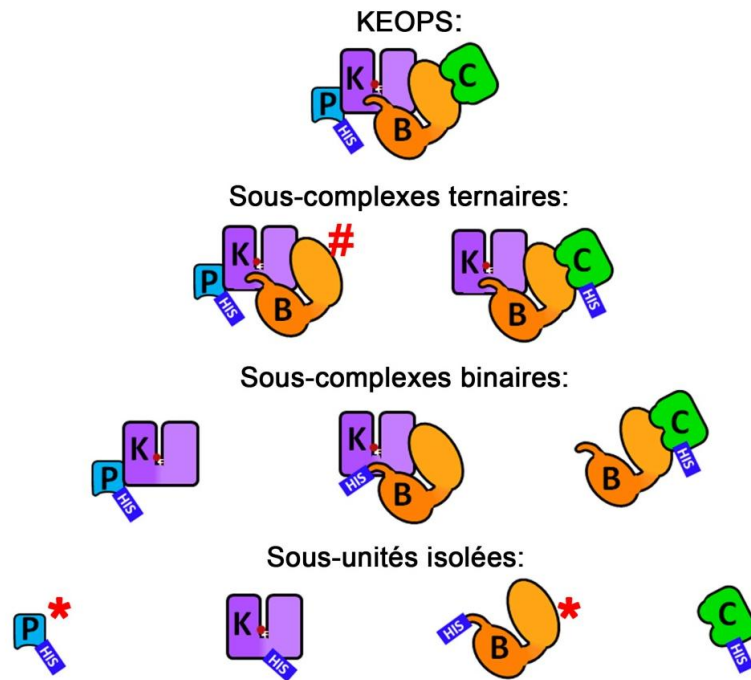
### C. Spectre d'absorption du complexe KEOPS purifié dans les mêmes conditions.

Les deux encarts correspondent aux mêmes spectres, mais avec des échelles de valeurs d'absorbance (en ordonnées) différentes : entre 0 et 1,2 UA pour l'encart de gauche, entre 0 et 0,1UA pour l'encart de droite.

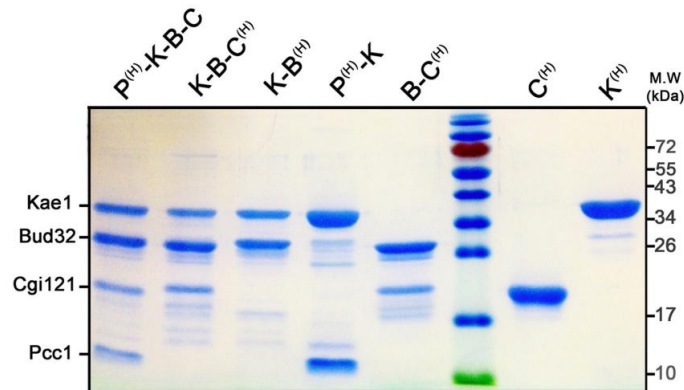
Concernant la protéine Cgi121<sup>-His</sup>, la majorité de la protéine purifiée à l'issue de cette première étape (FIGURE II-1-C, E) se maintient sous forme soluble et a été purifiée jusqu'à homogénéité grâce à une seconde étape de filtration sur gel. La protéine Cgi121<sup>-His</sup>, d'une masse moléculaire théorique de 19,8kDa, élue sous la forme d'un unique pic dont le volume d'élution correspond à une masse moléculaire apparente d'environ 25,9 kDa, pouvant correspondre à un état monomérique *in vitro* (FIGURE II-2-A). A l'issue de différentes purifications effectuées indépendamment, nous avons observé que la protéine Cgi121<sup>-His</sup> purifiée présente un spectre d'absorption avec un maxima centré sur 280nm accompagné dans certains cas d'un maxima secondaire centré autour de 305nm (qui forme un «épaulement» sur le pic principal). (FIGURE II-2-B). Le complexe KEOPS purifié dans les mêmes conditions présente un épaulement similaire, mais centré sur 340nm (FIGURE II-2-C). Bien que l'absorbance de la solution protéique de Cgi121<sup>-His</sup> purifiée soit relativement élevée à la longueur d'onde 260nm comparativement à 280nm (rapport 280/260= 1), aucun acide nucléique n'a pu être détectée par électrophorèse sur gel en condition dénaturante (non montré).

Les protéines Pcc1<sup>-His</sup> et Bud32<sup>-His</sup> précipitent totalement à l'issue de la chromatographie d'affinité sur résine de nickel (FIGURE II-1A-C, E), malgré différentes conditions testées (changement de position de l'étiquette histidine, ajout de détergents doux, augmentation de la force ionique, changement de pH). Cette précipitation survient extemporanément, au bout de quelques minutes après l'élution par l'imidazole, et est accélérée lorsque l'éluat est maintenu à 4°C. Ces précipitats, constitués respectivement des protéines Bud32<sup>-His</sup> et Pcc1<sup>-His</sup> pures et insolubles, ne contiennent pas d'acide nucléique, indiquant que ces sous-unités ne co-purifient pas avec des ARNt issus de la souche d'expression.

A.



B.



### FIGURE II-3 : Purification des sous-complexes de KEOPS

#### A. Représentation schématique des différentes combinaisons de sous-unités envisagées dans ce travail

La protéine Pcc1 (notée P) est représentée en bleu, la protéine Kae1 (notée K) en violet, la protéine Bud32 (notée B) en orange, la protéine Cgi121 (notée C) en vert. L'étiquette hexa-histidine est représentée par un rectangle bleu (notée HIS), dont la position schématique correspond à l'extrémité C-terminale de la protéine sur laquelle elle se trouve. L'étoile rouge indiquée à côté des sous-unités P-HIS et B-HIS indique que ces protéines ont pu être purifiées par chromatographie d'affinité mais sont instables en solution. Le symbole # indiqués à côté de la combinaison P-HIS K B indique que ce sous-complexe n'a pas pu être obtenu.

#### B. Analyse SDS-PAGE du complexe KEOPS (PKBC), des sous-complexes (KBC, KB, PK, BC) et des sous-unités (K, C) obtenus après purification par filtration sur gel.

Les différentes combinaisons de protéines ont été co-exprimées dans une souche Rosetta2(DE3)pLysS d'*Escherichia coli* à l'aide d'un milieu de culture auto-inducteur TB OverNightExpress+. Les complexes ont été ensuite purifiés par chromatographie d'affinité sur Ni-NTA, puis par gel filtration sur une colonne SuperDex 200 16/60 (selon Matériel et Méthodes). Chaque combinaison élue sous la forme d'un unique pic dont la fraction correspondante a été analysée par SDS-PAGE.

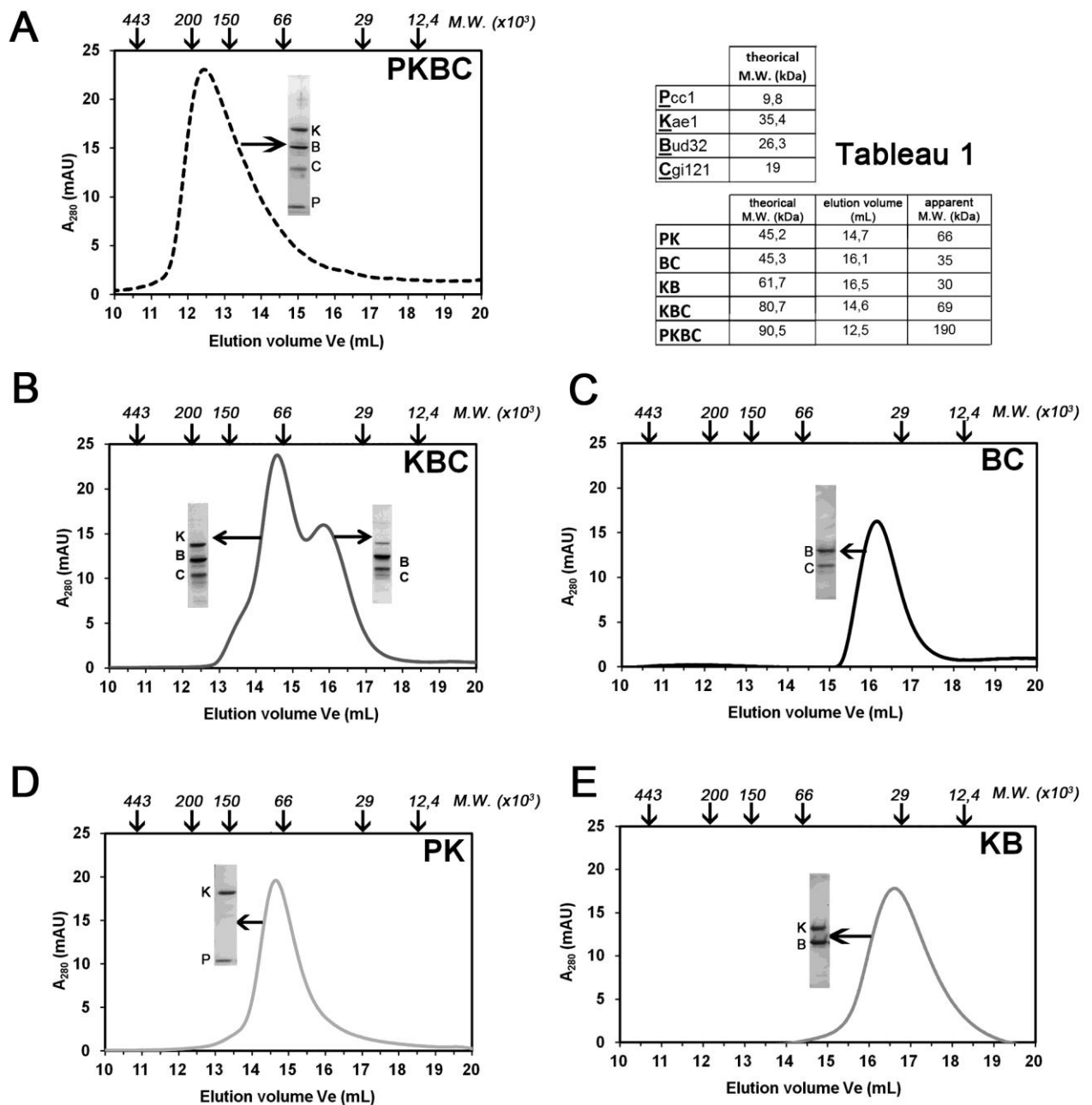
La composition des complexes est indiquée en haut du gel, l'étiquette hexahistidine est indiquée en exposant (H) à droite de la protéine sur laquelle elle est placée. La localisation des protéines au niveau du gel est indiquée à droite du gel et leur masse moléculaire théorique est indiquée entre parenthèse. Les masses moléculaires des marqueurs protéiques (puits central) sont indiqués sur la droite du gel en kilodalton (kDa).

## Purification des sous-complexes binaires et ternaires

Les protéines Bud32<sup>-His</sup> et Pcc1<sup>-His</sup> n'étant pas stables dans les différentes conditions *in vitro* testées, nous avons décidé de les co-exprimer avec leurs partenaires respectifs. D'après la caractérisation structurale du complexe KEOPS (Mao *et al.* 2008), les quatre protéines interagissent de façon linéaire deux à deux, selon l'ordre Pcc1-Kae1-Bud32-Cgi121. Pour plus de clarté, ces 4 sous-unités seront dorénavant appelées P, K, B et C ; le complexe KEOPS complet s'écrivant alors PKBC. Ainsi, seules certaines configurations sont susceptibles d'être obtenues par co-expression et co-purification : les sous-complexes PK, KB, BC, PKB et KBC (FIGURE II-3A).

Nous avons donc construit différents vecteurs polycistroniques d'expression, afin de produire ces sous-complexes. Dans chacune de ces associations, seule une protéine est porteuse d'une étiquette hexa-histidine en position C-terminale, indiquée par un symbole <sup>-H</sup>.

Les sous-complexes P<sup>-H</sup>K, KB<sup>-H</sup>, BC<sup>-H</sup> et KBC<sup>-H</sup> sont tous correctement exprimés dans *E.coli*, et se retrouvent à la fois dans les fractions insoluble et soluble après la lyse cellulaire (non montré). A l'issue d'une première étape de chromatographie d'affinité sur résine Ni-NTA, chacun des sous-complexes forme en partie un précipité insoluble qui apparaît extemporanément. Cependant, dans tous les cas, une large fraction de ces sous-complexes se maintient sous forme soluble qui a pu être purifiée jusqu'à homogénéité par filtration sur gel. Chacun des sous-complexes élue lors de cette étape sous la forme d'un unique pic, contenant les sous-unités attendues. Ces sous-complexes, tout comme le complexe P<sup>-H</sup>KBC et la protéine K<sup>-H</sup>, présentent une légère dégradation visible sous la forme de bandes supplémentaires en SDS-PAGE (FIGURE II-3B). Afin de s'assurer de leur stabilité et déterminer leur masse apparente, nous avons alors analysé un alicot de ces différents sous-complexes obtenus sur une colonne de filtration sur gel SuperDex-200 10/300 (FIGURE II-4). En référence, nous avons également analysé le complexe P<sup>-H</sup>KBC (FIGURE II-4A), et des marqueurs protéiques globulaires de taille connue.



**FIGURE II-4 : Analyse par filtration sur gel de la taille apparente des sous-complexes de KEOPS**

Un alicot d'environ 100µg de chaque complexe et sous complexe purifié a été décongelé sur glace et analysé sur une colonne Super200 10/300<sup>(GE)</sup>, à 4°C, dans le même tampon que celui utilisé pour la purification. La quantité de protéine en sortie de colonne a été suivie par mesure de l'absorbance à 280nm (en ordonnée), et les fractions d'élution ont été analysées par SDS PAGE (la fraction correspondant au maxima est montrée à côté de chaque pic). Le volume d'élution de différents marqueurs protéiques globulaires de poids moléculaire connu (M.W.) sont indiqués en kDa par des flèches sur la partie supérieure de chaque chromatogramme. (apoprotinine 443,  $\beta$ -amylase 200, alcool deshydrogenase 150, albumine 66, carbonic anhydrase 29, cytochrome C 12,4 kDa) La masse moléculaire théorique, le volume d'élution et la masse moléculaire apparente des sous-unités et sous-complexes sont indiqués dans le tableau 1.

**A. Chromatogramme obtenu avec le complexe P<sup>H</sup>KBC (=KEOPS).**

**B. Chromatogramme obtenu avec le complexe KBC<sup>H</sup>.**

**C. Chromatogramme obtenu avec le complexe BC<sup>H</sup>.**

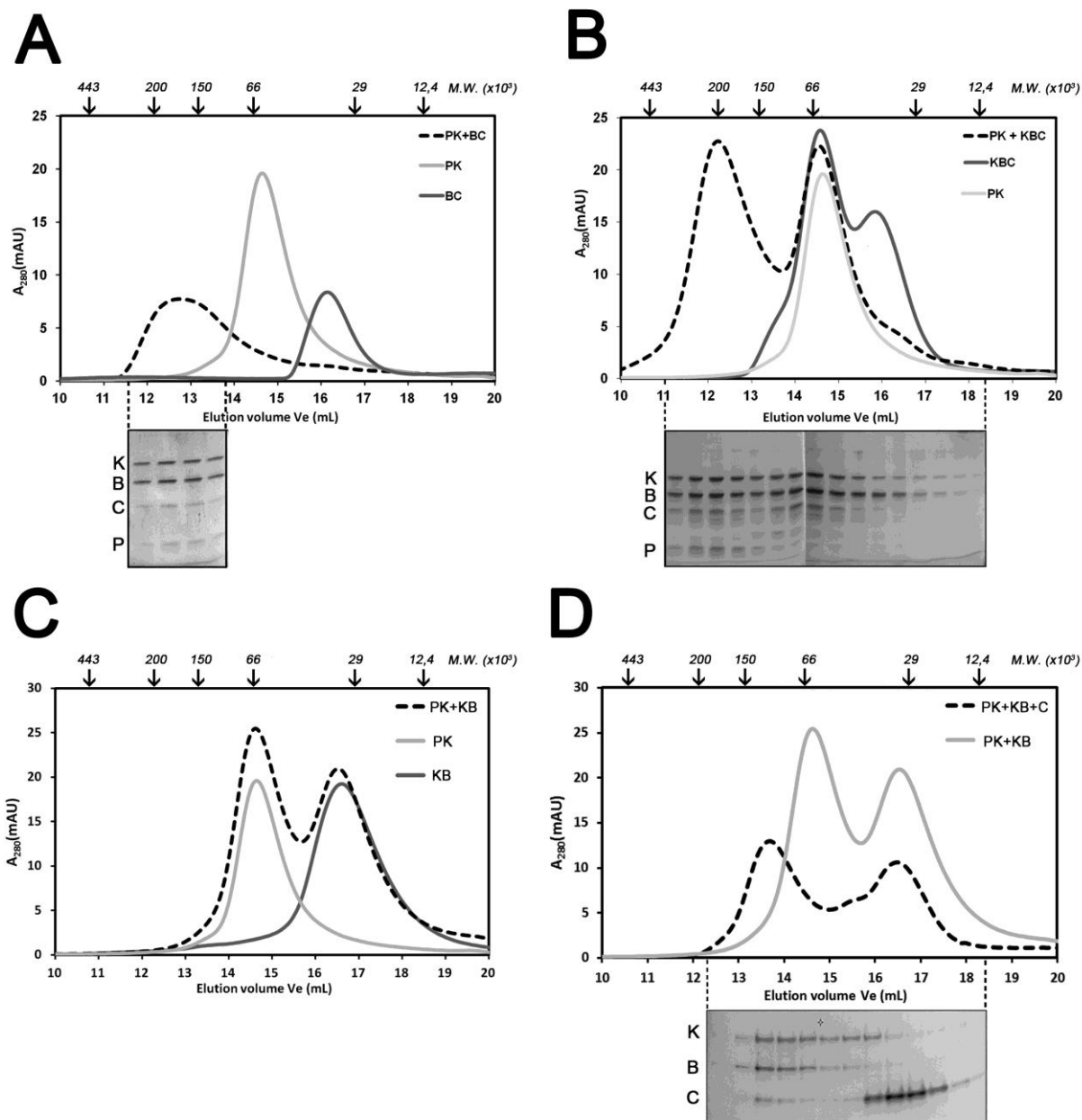
**D. Chromatogramme obtenu avec le complexe P<sup>H</sup>K.**

**E. Chromatogramme obtenu avec le complexe KB<sup>H</sup>.**



Le complexe  $P^H K$  (masse moléculaire théorique = 45 kDa) est stable et se maintient sous la forme d'un pic unique correspondant à une masse moléculaire apparente de 66 kDa ; pouvant correspondre à un dimère  $P^H K/P^H K$  (FIGURE II-4B). Les complexes  $KB^H$ ,  $BC^H$  sont également stables en solution et présentent respectivement une masse moléculaire apparente de 30 kDa, 35 kDa, pouvant correspondre à leur forme monomérique respective (FIGURE II-4C, D). Le sous-complexe  $KBC^H$ , qui a été isolé sous la forme d'un unique pic d'élution lors de sa purification, montre l'apparition de deux pics lors de cette analyse : un pic majoritaire, contenant les trois protéines K, B, et  $C^H$  en quantité équimolaire, d'une masse apparente de 69 kDa, correspondant à la forme monomérique du complexe ternaire ; accompagné d'un pic secondaire, contenant les protéines B et  $C^H$ . Cela indique que le sous complexe  $KBC^H$  se dissocie spontanément dans une certaine mesure (FIGURE II-4E).

A l'issue de ces purifications, les sous-complexes  $P^H K$ ,  $KB^H$  et  $KBC^H$  présentent un maxima d'absorption secondaire centré à 500nm, caractéristique de la présence du fer dans la protéine K (non montré). En revanche, les sous complexe contenant la protéine C ( $BC^H$  et  $KBC^H$ ) ne montrent pas l'épaule à 305nm observé lors de la purification de  $C^H$  seule. Enfin, comme pour les sous-unités isolées, nous n'avons détecté aucun acide nucléique co-purifiant avec ces différents sous-complexes.



**FIGURE II-5 : Test d'interaction entre les sous-complexes de KEOPS par gel filtration**

Différentes combinaisons de sous-complexes (indiquées par la suite) ont été réalisées, en mélangeant environ 100µg de chaque sous-complexe dans un volume final de 200µL, dans le tampon utilisé pour leur purification. Après une incubation de 15min à 20°C (température ambiante), les mélanges ont été séparés sur une colonne de filtration sur gel SuperDex200 10/300, à 4°C. La quantité de protéine en sortie de colonne a été suivie par mesure de l'absorbance à 280nm (en ordonnée), et les fractions d'élution ont été analysées par SDS-PAGE pour déterminer leur composition (non montrée). Le volume d'élution des marqueurs globulaires est indiqué en haut de chaque graphe (voir Figure II-4). La courbe en pointillé est le chromatogramme correspondant au mélange, les courbes en traits pleins correspondent aux chromatogrammes des sous-complexes seuls (ou mélange pour la partie D.). Dans le mélange PK+KB+C (D.), un excès de protéine C<sup>H</sup> (300µg) a été ajouté au mélange P<sup>H</sup>K+KB<sup>H</sup>, et le flux de tampon en FPLC a été réduit. L'analyse SDS PAGE correspondant à certaines fractions d'élution est montrée sous chaque graphe.

- A.** Combinaison de P<sup>H</sup>K + BC<sup>H</sup> **B.** Combinaison de P<sup>H</sup>K + KBC<sup>H</sup>  
**C.** Combinaison de P<sup>H</sup>K + KB<sup>H</sup> **D.** Combinaison de P<sup>H</sup>K + KB<sup>H</sup> + C<sup>H</sup>

A l'issue des purifications décrites précédemment, nous disposons au final des sous-complexes et sous-unités suivants :  $P^H K$ ,  $K^H$ ,  $KB^H$ ,  $BC^H$ ,  $KBC^H$  et  $C^H$ , et à titre de référence le complexe KEOPS complet  $P^H KBC$  (voir **FIGURE II-3B**)

## Reconstitution physique du complexe KEOPS

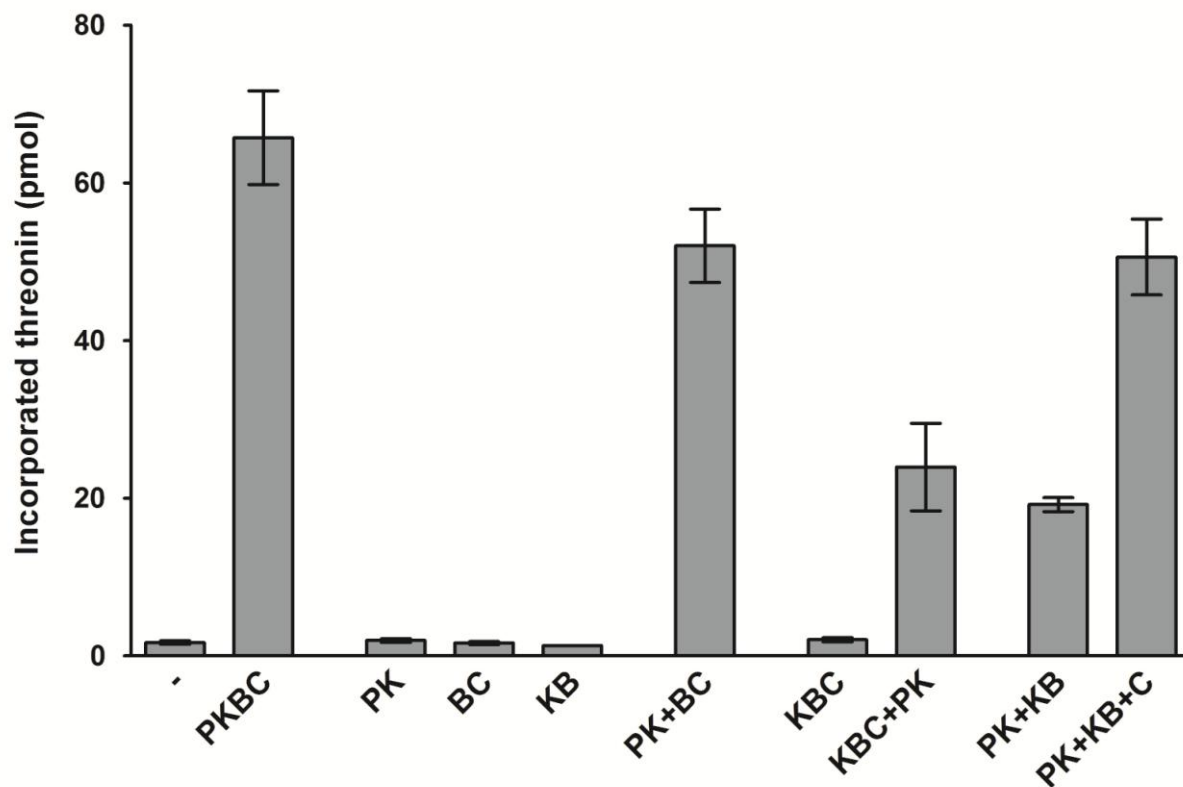
Dans un premier temps, nous avons vérifié que les sous-complexes produits peuvent reformer un complexe complet PKBC par étude interactionnelle. Différentes combinaisons de sous-complexes ont été incubées pendant 30 min à 20°C dans le tampon utilisé pour la purification, puis le mélange a été séparé par filtration sur gel à 4°C sur une colonne SuperDex200-10/300 (**FIGURE II-5**). Les fractions d'élution obtenues ont ensuite été analysées par SDS PAGE pour vérifier la composition des pics d'élution obtenus.

Un mélange équimolaire de  $[P^H K + BC^H]$  élue sous la forme d'un unique complexe à un volume de 12,7mL, semblable au volume d'élution du complexe  $P^H KBC$  à 12,5mL, et contenant les 4 sous-unités (visibles en SDS PAGE) indiquant la reconstitution complète du complexe KEOPS (**FIGURE II-5A**)

Un mélange de  $[P^H K + KBC^H]$  provoque également l'apparition d'un pic semblable contenant les 4 sous unités, indiquant également la reconstitution d'un complexe KEOPS complet (**FIGURE II-5B**)

En revanche, le chromatogramme correspondant au mélange  $[P^H K + KB^H]$  est identique à la superposition des chromatogrammes issus des sous complexes isolés, indiquant qu'un complexe PKB stable n'a pas été formé dans ces conditions (**FIGURE II-5C**).

Afin de savoir si la protéine C peut permettre de stabiliser l'association PKB, nous avons effectué un mélange identique, en ajoutant un excès de protéine C (**FIGURE II-5D**). Dans ce cas, le chromatogramme obtenu est relativement complexe, mais on constate l'apparition d'un nouveau pic contenant les 4 sous-unités (visible en SDS PAGE). Le pic intermédiaire à 15,4mL contient majoritairement la protéine K et le dernier pic à 16,5mL contient majoritairement la protéine  $C^H$ . Ces données montrent qu'en présence de  $C^H$ , les sous complexe  $P^H K$  et  $KB^H$  ont pu s'associer pour former un complexe reconstitué  $P^H KB^H C^H$ , associé à une libération de K en excès. Il est à noter que dans cette expérience, pour des raisons techniques, le flux de tampon a été réduit, ce qui peut amener à des décalages du volume d'élution observé pour le complexe complet (13,6mL contre 12,7mL).



**FIGURE II-6 : Activité t<sup>6</sup>A des différents sous-complexes de KEOPS**

Les différents sous-complexes (1μM) ont été ajoutés seuls ou en combinaison à un mélange réactionnel contenant : 2μM de la protéine Sua5<sup>-His</sup>, 18,2μM de <sup>14</sup>C-L-thréonine, 1mM d'ATP, 5mM de carbonate, 100μM de MnCl<sub>2</sub>, et XμM d'ARN<sup>Lys(UUU)</sup> d'*E.coli* produit par sur-expression *in vivo*, dans un tampon HEPES 50mM, pH8, KCl 35mM, DTT1mM, 5mM de MgCl<sub>2</sub>. Après 30min. d'incubation à 55°C, le mélange réactionnel a été précipité au TCA, et la quantité de <sup>14</sup>C-L-thréonine incorporée a été mesurée par scintillation (en ordonnée). Les différentes combinaisons de sous-complexes sont indiquées en abscisse. - : mélange réactionnel seul. Les barres d'erreurs indiquent la déviation standard de la moyenne obtenue par trois expériences indépendantes.

## Les quatres sous-unités sont requises pour la biosynthèse optimale de t<sup>6</sup>A

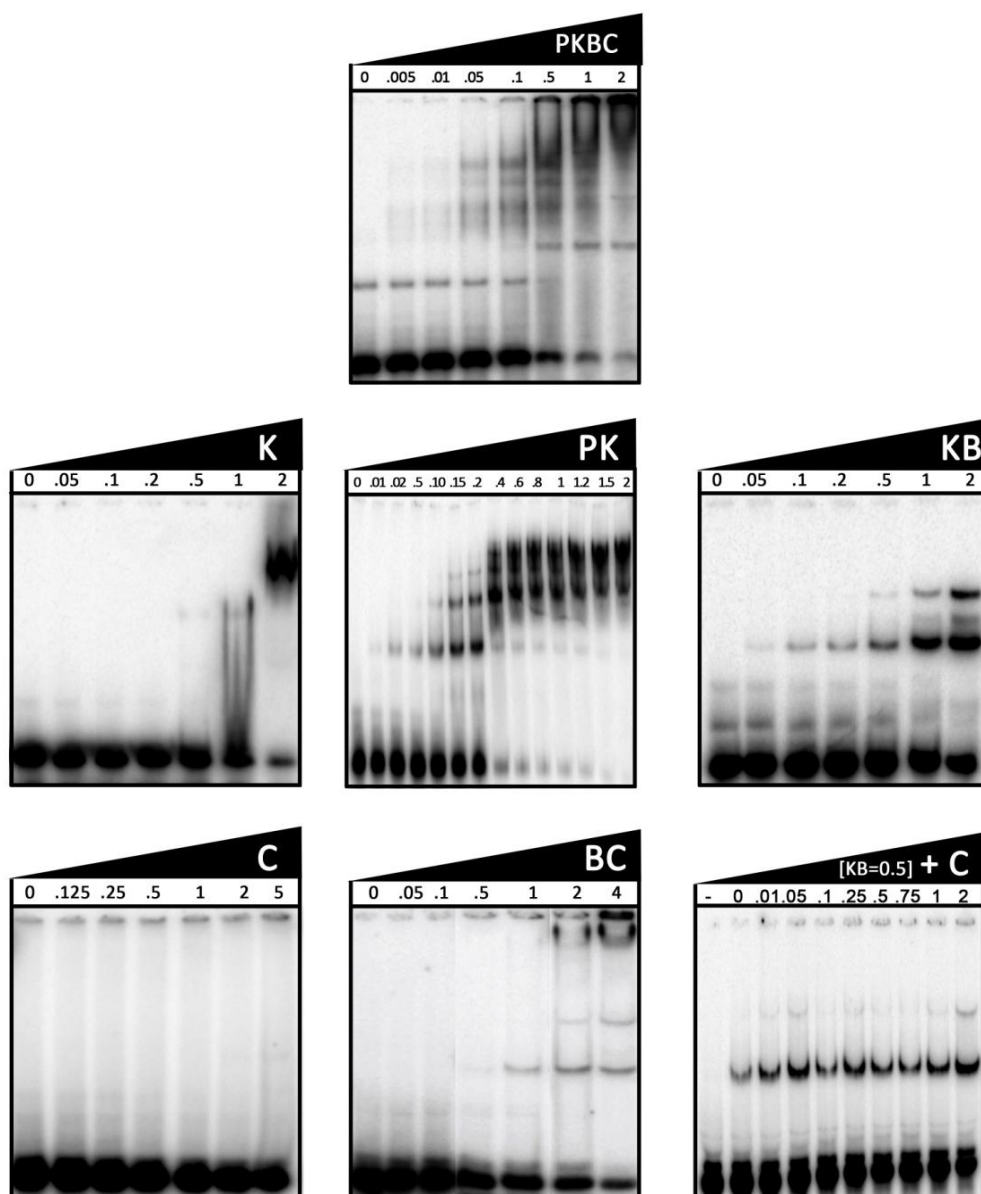
Les différents sous-complexes, seuls ou combinés de manière équimolaire, ont ensuite été testés pour leur capacité à catalyser la formation de t<sup>6</sup>A selon le protocole décrit précédemment (PERROCHIA ET AL 2013, MATERIEL ET METHODES), en présence de la protéine Sua5 de *P.abysyi*, d'ARNt substrat, d'ATP, de Mn<sup>2+</sup> et Mg<sup>2+</sup>, de carbonate et de C<sup>14</sup>L-thréonine (FIGURE II-6). On appellera « activité t<sup>6</sup>A » l'incorporation de C<sup>14</sup>L- thréonine radiomarquée au niveau des ARNt, traduisant la formation du nucléoside modifié t<sup>6</sup>A.

Les sous-complexes P<sup>-H</sup>K ou BC<sup>-H</sup> seuls ne présentent pas d'activité t<sup>6</sup>A, alors que leur combinaison restaure la majeure partie (80%) de l'activité t<sup>6</sup>A observée avec le complexe P<sup>H</sup>KBC

De même, le complexe KB<sup>-H</sup> (absence de P et de C dans le milieu réactionnel) ne présente aucune activité t<sup>6</sup>A, suggérant que P et/ou C sont nécessaires à cette activité.

Le sous-complexe KBC<sup>-H</sup> ne présente aucune activité détectable, mais l'addition du complexe P<sup>-H</sup>K permet de restaurer partiellement (~36%) cette activité. De plus, la combinaison [P<sup>-H</sup>K + KB<sup>-H</sup>], où C est absent du milieu réactionnel, conduit à une activité t<sup>6</sup>A modérée (~29%). Ces résultats indiquent que P, K et B sont nécessaires et suffisants à l'activité de biosynthèse de t<sup>6</sup>A.

L'ajout en excès de la protéine C<sup>-H</sup> au mélange [P<sup>-H</sup>K + KB<sup>-H</sup>], augmente de 2 fois la synthèse de t<sup>6</sup>A, restaurant une activité t<sup>6</sup>A similaire à celle du complexe entier P<sup>-H</sup>KBC. Ainsi, nos données indiquent que C ne n'est pas strictement nécessaire à une activité t<sup>6</sup>A mais est requise pour une activité optimale.



**FIGURE II-7 : Propriétés de fixation à l'ARNt par les sous-complexes de KEOPS en retard du gel**

Une concentration fixe (10nM) d'ARNt<sup>Lys(UUU)</sup> *Pyrococcus abyssi* transcrit *in vitro* et radiomarké, a été incubée à température ambiante pendant 15mn en présence de différentes concentrations de sous-complexes, dans un tampon TrisHCl 50mM, pH8, NaCl 100mM, MgCl<sub>2</sub> 5mM, DTT 1mM. Les complexes nucléo-protéiques ont été séparés par électrophorèse sur gel d'acrylamide (19 :1) 6% en condition native à température ambiante. La radioactivité présente dans le gel a été visualisée par phosphorimaging. Le type de sous-complexe analysé et la concentration en protéine dans chaque puits sont indiqués sur la partie supérieure de chaque cadran. La position des puits du gel et de l'ARNt libre correspondent respectivement aux parties supérieure et inférieure de chaque cadran.

Dans le cadran inférieur droit, une concentration fixe de sous-complexe KB<sup>H</sup> (0,5μM) a été additionnée de la protéine C<sup>H</sup> à différentes concentrations. — : absence de KB et C.

## Le couple Pcc1-Kae1 est le principal élément de fixation aux ARNt

Précédemment (PERROCHIA ET AL., 2013), nous avons étudié les propriétés de fixation *in vitro* du complexe KEOPS ( $P^H$ KBC) aux ARNt par la technique de retard sur gel. Nous avons montré que la présence de ce complexe provoque un retard de migration des ARNt radiomarqués sur gel d'acrylamide en condition native, détectable dès 10nM de protéine, (correspondant dans nos essais à un ratio ARNt/protéine 1 :1), et visible sous la forme de multiple bandes retardées correspondant à différentes configurations de complexes nucléo-protéiques (FIGURE II-7, ENCART SUPERIEUR). A partir de 500nM, 50% de l'ARNt substrat sont retardés, et au-delà de 2 $\mu$ M, la totalité de l'ARNt est retardé.

Nous avons donc employé cette technique dans les mêmes conditions afin de déterminer l'affinité relative des différents sous-complexes de KEOPS pour les ARNt *in vitro*, et savoir quels composants étaient responsables de cette fixation (FIGURE II-7).

La protéine  $K^H$  seule provoque un retard de migration détectable des ARNt à partir de 500nM. Une concentration de 1 $\mu$ M correspond à une zone de transition entre une forme libre et une forme complexé ARNt/protéine. La quasi-totalité de la sonde ARNt est mobilisée au-delà de cette concentration, sous la forme d'une bande diffuse migrant dans le gel. Cette zone de transition entre forme libre et forme complexé de l'ARNt est relativement restreinte comparativement au profil obtenu avec le complexe  $P^H$ KBC, et correspond à un phénomène de « tout ou rien ».

Le complexe  $P^H$ K provoque un retard de migration de l'ARNt se traduisant par l'apparition progressive de plusieurs bandes discrètes au sein du gel, détectable à partir de 10nM de protéine, de manière similaire au complexe  $P^H$ KBC entier. Entre 200nM et 400nM de protéine, on observe une transition correspondant à la fixation de la majorité de l'ARNt sous la forme d'une nouvelle bande semblable à celle observée pour la protéine  $K^H$  seule. Le comportement de ce sous-complexe traduit donc une forte augmentation de l'affinité pour les ARNt par rapport à la protéine  $K^H$  seule, et se rapproche du comportement du complexe  $P^H$ KBC.



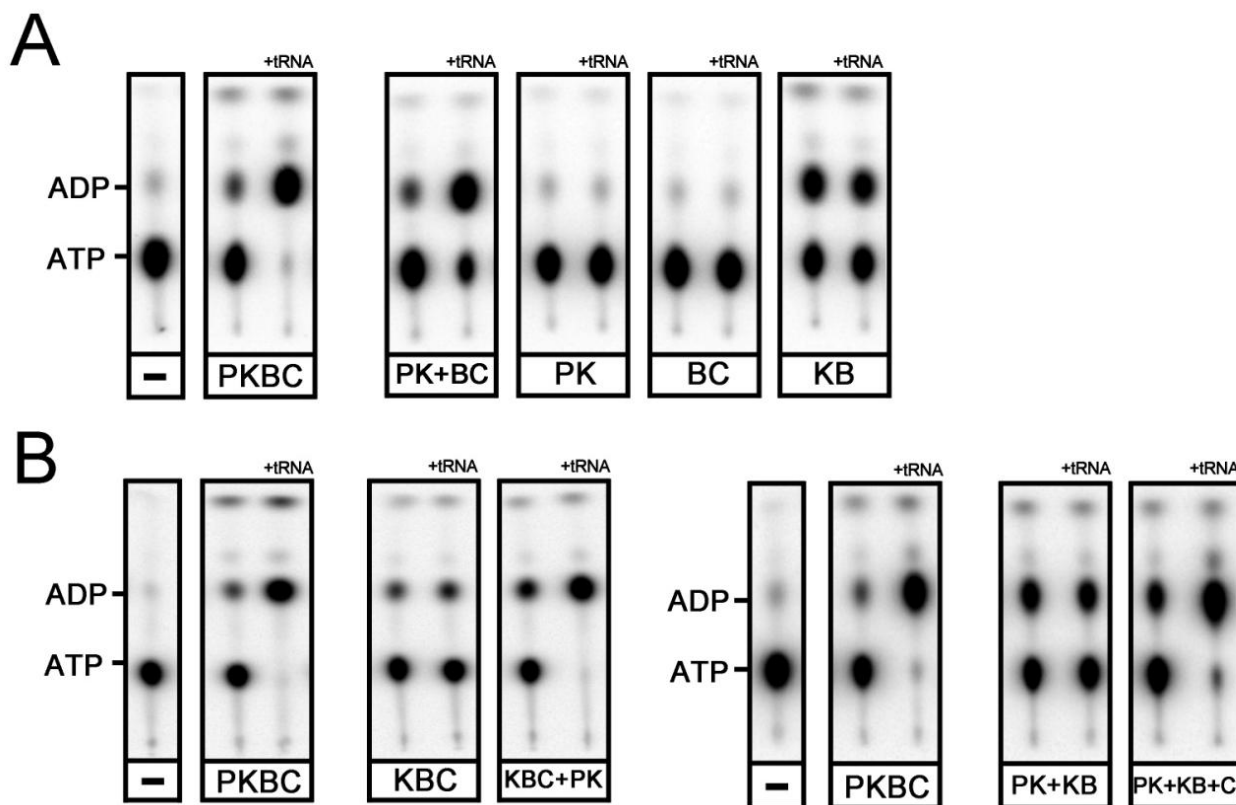


Le complexe  $KB^{-H}$  provoque également l'apparition progressive de deux bandes discrètes retardées, détectable à partir de 50nM de protéines. L'intensité de ce retard est cependant moins forte que pour le complexe PK : environ 50% seulement des ARNt sont retardés à une concentration de 2 $\mu$ M, concentration maximale testée pour ce sous-complexe. La présence de la protéine B modifie donc le comportement de K, mais de manière moins importante que la protéine P.

Le complexe  $BC^{-H}$  provoque en revanche un retard beaucoup moins prononcé. Deux bandes retardées sont observés à 1 $\mu$ M de protéine. A plus forte concentration en protéine, seule une faible fraction de l'ARNt est retardée et on constate l'apparition de complexe nucléo-protéique au niveau des puits du gel, correspondant à un phénomène d'agrégation.

Enfin, la protéine  $C^{-H}$  seule ne provoque aucun retard de migration apparent, même à forte concentration protéique. Etant donné l'effet stimulateur que présente  $C^{-H}$  vis-à-vis de l'activité  $t^6A$  montré précédemment, nous avons donc cherché à savoir si cette protéine pouvait provoquer une augmentation de l'affinité du complexe  $KB^{-H}$  pour les ARNt. Nous avons donc incubé l'ARNt en présence d'une concentration fixe du sous-complexe  $KB^{-H}$ , et en ajoutant la protéine  $C^{-H}$  en concentration croissante. Cet ajout ne provoque aucun changement dans le profil de migration.

Ces différents résultats indiquent que K et P sont les protéines principales responsables de la fixation des ARNt, la protéine B pouvant également intervenir dans cette fixation.



### FIGURE II-8 : Activité ATPase des sous complexes de KEOPS

Les différents sous-complexes (1 $\mu$ M) ont été incubés dans le même tampon utilisé précédemment (Figure-II6), en présence de 10 $\mu$ M d'ATP supplémenté avec 1 $\mu$ Ci d'ATP<sup>[ $\alpha$ P<sup>32</sup>]</sup>, et en l'absence ou en présence d'ARNt<sup>Lys(UUU)</sup> (E.co) 2 $\mu$ M (indiqué par +tRNA en haut de chaque colonne). Après 30min. d'incubation à 55°C, les produits de l'hydrolyse de l'ATP ont été séparés par chromatographie sur couche mince (CCM) et les composés radioactifs ont été visualisés par exposition sur un écran au phosphore et révéler à 533nm (« phosphorimaging »).. Les différentes combinaisons de sous-complexes ajoutées au mélange réactionnel sont indiquées sous chaque colonne. Le complexe PKBC a été utilisé dans chaque expérience comme référence. La localisation de l'ADP et de l'ATP est indiquée pour la condition -, correspondant au mélange réactionnel sans protéine.

**A. Activité des sous-complexes à deux composants : PK, BC et KB.**

**B. Activité des sous-complexes et combinaisons à trois composants : KBC, PK+KB.**

**C. Activité des sous-complexes mutants.** Les protéines K et B ont été mutées ponctuellement au niveau d'un résidu, et purifiées sous la forme de sous-complexes PK et BC mutants, respectivement notés PK\*, PK\* et B\*<sup>C</sup>. L'activité ATPase de ces sous-complexes a été suivie selon le protocole décrit ci-dessus. Le nature et la position de la mutation est indiquée sur la droite de la figure, ainsi que la correspondance des symboles. D=acide aspartique, A= alanine, H= histidine.

## Identification des éléments responsables de l'activité ATPase

Précédemment, nous avons montré que le complexe KEOPS, dans les conditions *in vitro* de biosynthèse de t<sup>6</sup>A, présente une activité d'hydrolyse de l'ATP en ADP + Pi fortement stimulée par la présence d'ARNt. Afin de comprendre le rôle des composants de KEOPS dans cette activité, nous avons suivi par chromatographie sur couche mince la production d'ADP à partir d'ATP radiomarké au niveau du phosphate  $\alpha$  par les différents sous-complexes, en absence ou en présence d'ARNt.

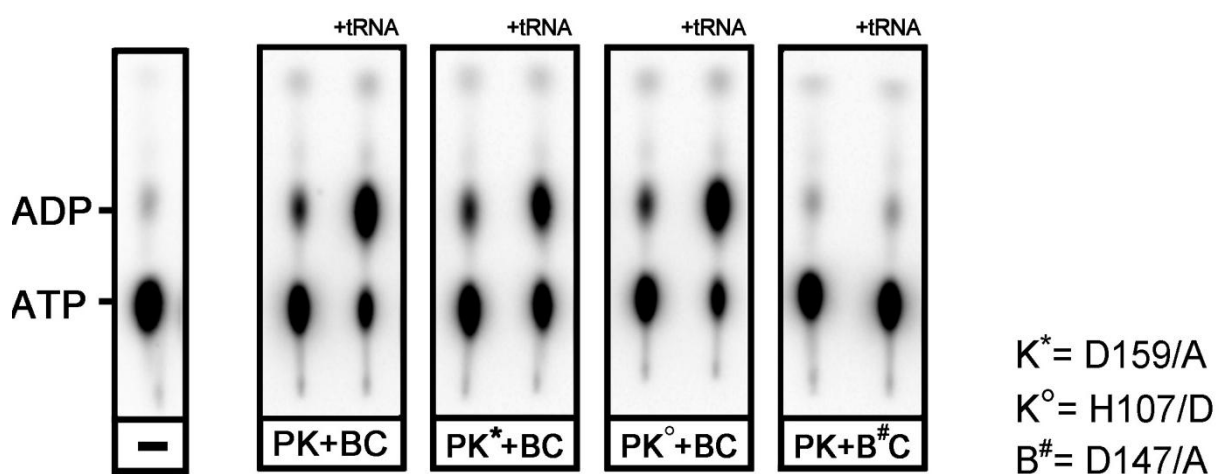
### **Le couple Kae1-Bud32 est responsable de l'activité ATPase de KEOPS**

Les sous complexes binaires P<sup>H</sup>K et BC<sup>H</sup> présentent chacun une très faible activité ATPase, en l'absence ou en présence d'ARNt. En revanche, le sous complexe binaire KB<sup>H</sup> présente une activité ATPase comparable à celle du complexe P<sup>H</sup> KBC, mais qui n'est pas stimulée par la présence d'ARNt (**FIGURE II-8A**). Cela indique que K et/ou B sont responsables de cette activité, et que la présence de P et/ou C permet une stimulation par les ARNt.

### **Les protéines Pcc1 et Cgi121 sont nécessaires à la stimulation par les ARNt**

Le sous complexe KBC<sup>H</sup> présente une activité ATPase comparable à celle du complexe KB<sup>H</sup>, qui est insensible à la présence d'ARNt. L'ajout du sous complexe P<sup>H</sup>K permet de restaurer la réponse à l'ARNt. De même, la combinaison P<sup>H</sup>K+KB<sup>H</sup> présente une activité ATPase comparable à celle du complexe KB<sup>H</sup>; et est insensible à la présence d'ARNt, alors que l'addition de la protéine C<sup>H</sup> permet d'obtenir une réponse à l'ARNt (**FIGURE II-8B**).

Ainsi, ces données montrent que la présence simultanée de C et P est nécessaire pour obtenir une réponse à l'ARNt.



**FIGURE II-9 : Activité ATPase des sous complexes mutants**

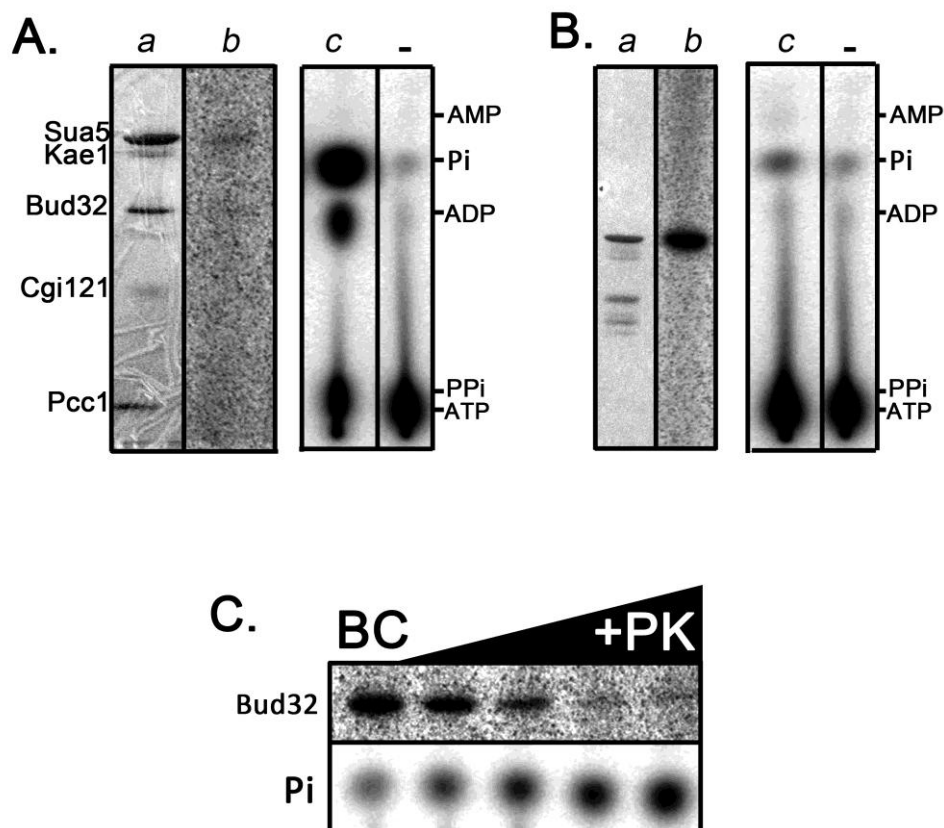
Les protéines K et B ont été mutées ponctuellement au niveau d'un résidu, et purifiées sous la forme de sous-complexes PK et BC mutants, respectivement notés PK\*, PK° et B<sup>#</sup>C. L'activité ATPase de ces sous-complexes a été suivie selon le protocole décrit dans la figure II-9. Le nature et la position de la mutation est indiquée sur la droite de la figure, ainsi que la correspondance des symboles. D=acide aspartique, A= alanine, H= histidine.

### **La protéine Bud32 est l'ATPase du complexe KEOPS**

Les données structurales indiquent que K et B contiennent chacune un site de fixation au nucléotide (HECKER ET AL, 2007 ; HECKER ET AL, 2008). Afin de déterminer qui de K ou de B est responsable de l'activité l'ATPase du complexe, nous avons donc muté ponctuellement ces protéines au niveau de leur site actif respectif. Pour la protéine B, nous avons remplacé l'acide aspartique conservé (D147) présent au niveau de la boucle catalytique par une alanine. Pour la protéine K, nous avons touché deux positions strictement conservées chez cette famille de protéines: d'une part, l'histidine (H107) impliquée dans la coordination du fer, et d'autre part, l'acide aspartique (D159) impliqué dans la liaison au nucléotide. Ces protéines mutantes ont été produits respectivement sous la forme de sous-complexe  $BC^{-H}$  et  $P^{-H}K$  (notés  $B_{D/A}C^{-H}$ ,  $P^{-H}K_{D/A}$  et  $P^{-H}K_{H/A}$ ). Le sous-complexe  $P^{-H}K_{H/A}$  ne présente plus d'absorption à 500nm associée à une couleur rose de la solution protéique, suggérant la perte de l'atome de fer au niveau du site actif (non montré).

Les combinaisons  $P^{-H}K_{D/A} + BC^{-H}$  et  $P^{-H}K_{H/A} + BC^{-H}$  présentent une activité ATPase similaire au mélange sauvage PK+BC (c'est-à-dire une hydrolyse de l'ATP stimulée par la présence d'ARNt). En revanche, le mélange  $P^{-H}K + B_{D/A}C^{-H}$  ne présente aucune activité ATPase (FIGURE II-9).

Ces résultats indiquent que la perturbation du site actif de K n'a pas d'impact sur l'activité ATPase du complexe KEOPS, et que la protéine B est responsable de cette activité.



**FIGURE II-10 : Activité phosphotransférase de B dans différentes conditions**

Un mélange réactionnel contenant les différents composants indiqués ci-dessous a été incubé dans le tampon utilisé dans la Figure II-6 en présence d'ATP<sup>[<sup>32</sup>P]</sup> (1μCi) pendant 30mn à 55°C. Un alicot de la réaction (1/20<sup>ème</sup>) a été analysé par CCM comme indiqué dans la Figure II-7 (c), et le reste du mélange réactionnel a été analysé par SDS-PAGE. Les protéines ont été révélées par coloration au bleu de Coomassie (a), et la radioactivité a été visualisée par phosphorimaging (b). La localisation des protéines en SDS PAGE est indiquée sur la gauche du cadran a, la localisation des produits d'hydrolyse de l'ATP en CCM est localisé sur la droite des cadrans c. - = mélange réactionnel analysé par CCM en l'absence de protéine.

**A. Analyse de l'activité du complexe PKBC en présence des composants de la biosynthèse de t<sup>6</sup>A.**

Le mélange réaction contient : P-HKBC (1,5μM) + Sua5 (3μM) + ARNt (12μM) + carbonate (5mM) + thréonine (100μM)

**B. Analyse de l'activité du sous complexe BC.**

Le mélange réactionnel contient : BC (1,5μM)

**C. Analyse de l'activité du sous complexe BC en présence du sous complexe PK.**

Le mélange réactionnel contient une concentration fixe de BC (1,5μM) additionnée d'une concentration croissante de PK (0, 0.3, 0.6, 1.2, 1.6μM). Seule la zone correspondant à la protéine B en SDS PAGE, et la zone correspondant au Pi en CCM, sont montrées.

### **Kae1 inhibe l'activité Phosphotransférase de Bud32**

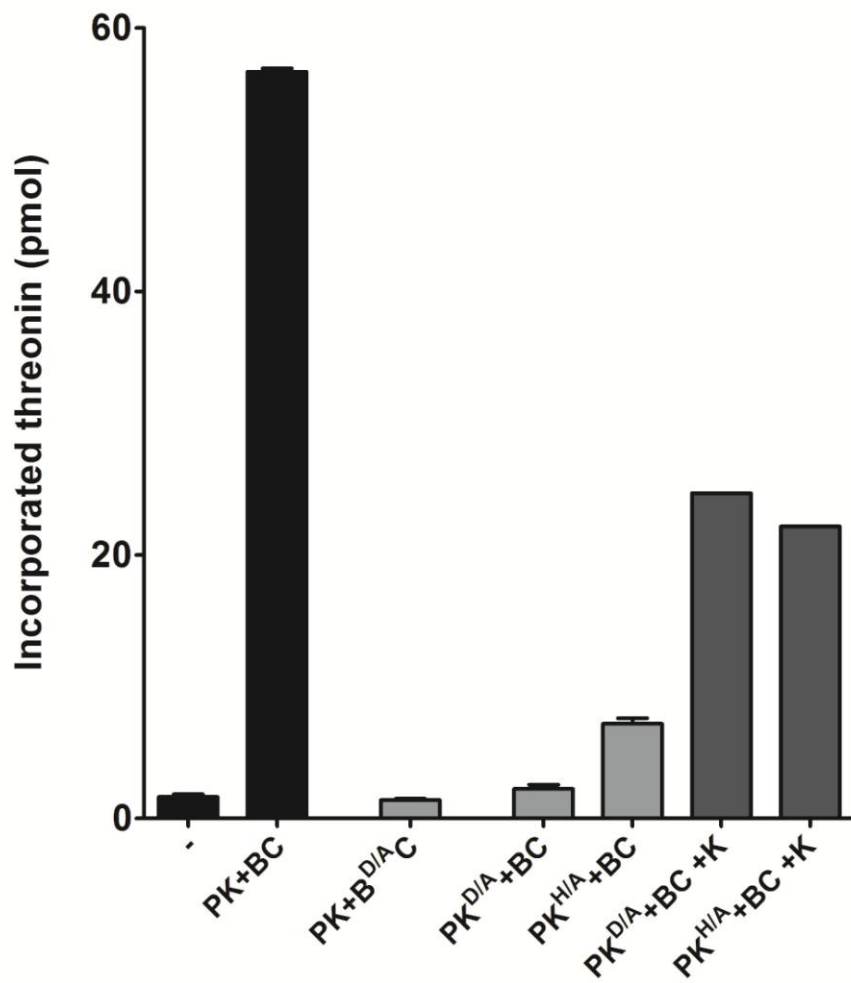
Précédemment (résultat complémentaire, Partie I), nous avons montré que le phosphate produit par l'hydrolyse de l'ATP par le complexe KEOPS est libéré en majorité sous forme de phosphate (Pi) libre en solution. De plus, nous venons de montrer que la protéine B est responsable de cette activité. Or, les orthologues de cette protéine chez *M. jannaschii* et *S. cerevisiae* ont été précédemment décrite comme des protéines kinases présentant une activité d'autophosphorylation *in vitro* (FACCHIN ET AL, 2002, HECKER ET AL, 2008, MAO ET AL, 2008). Nous avons donc voulu savoir si l'hydrolyse de l'ATP provoqué par B pouvait correspondre à un transfert de phosphate sur l'une des protéines du système.

Dans un premier temps, nous avons donc incubé le complexe  $P^H$ KBC en présence d'ATP radiomarké au niveau du phosphate gamma dans les conditions de synthèse de  $t^6A$ , c'est à dire en présence des différents composants de la réaction, puis avons suivi la radioactivité présente au niveau des protéines après séparation sur gel SDS-PAGE. Aucune phosphorylation significative n'est détectable, alors qu'une forte production de Pi libre est obtenue. (FIGURE II-10A). Ainsi, dans le contexte KEOPS, la protéine B ne présente pas d'activité phosphotransférase mais libère le phosphate inorganique sous forme libre suite à l'hydrolyse de l'ATP.

Nous avons alors incubé le complexe  $BC^H$  seul, dans des conditions similaires. Dans ce cas, la protéine B est autophosphorylée, contrairement à la protéine B mutée au niveau de son site actif (FIGURE II-10B). Ces données indiquent que la protéine B, en présence de C, présente une activité phosphotransférase qui implique le même site actif que l'activité ATPase décrite précédemment.

Enfin, nous avons montré que l'addition d'une concentration croissante du sous complexe  $P^H$ K au sous-complexe  $BC^H$  s'accompagne d'une diminution de l'autophosphorylation observée, corrélée à une augmentation de la quantité de Pi libre (FIGURE II-10C)

Ces données suggèrent donc que la présence de K module l'activité ATPase de B, en inhibant l'activité de phosphotransférase et en favorisant la libération de phosphate libre.



**FIGURE II-11 : Activité t<sup>6</sup>A des sous-complexes mutants de KEOPS**

L'activité de biosynthèse de t<sup>6</sup>A des sous-complexes mutants PK<sup>\*</sup>=PK<sub>H/A</sub>, PK<sup>\*</sup>=PK<sub>D/A</sub>, et B#C=B<sub>D/A</sub>C décrits dans la Figure II-7 C a été suivie dans les conditions décrites dans la Figure II-6, en présence respective des complexes non mutants BC ou PK. La présence des différents sous-complexes est indiquée sous le graphique. Dans les deux dernière conditions, les mélange PK<sup>\*</sup>+BC et PK<sup>\*</sup>+BC ont été additionné d'une quantité équimolaire de la protéine K<sup>H</sup>. - : incorporation en présence de Sua5 uniquement. La combinaison PK+BC a été utilisée comme référence



### **L'activité ATPase de Bud32 est nécessaire à la biosynthèse de t<sup>6</sup>A**

Dans un dernier temps, nous avons voulu savoir si l'activité ATPase-ARNt-dépendante de B joue un rôle dans la biosynthèse de t<sup>6</sup>A. Aucune production de t<sup>6</sup>A n'est observée avec la combinaison  $P^H K + B_{D/A} C^H$ , indiquant que cette hydrolyse d'ATP est indispensable à la synthèse de t<sup>6</sup>A (FIGURE II-11).

### **La fixation au nucléotide par Kae1 et la présence du Fer sont essentiels à l'activité de biosynthèse de t<sup>6</sup>A**

La perturbation du site actif de K par mutation n'a aucun impact sur l'activité ATPase décrite précédemment. De plus, les sous-complexes mutants  $P^H K_{D/A}$  et  $P^H K_{H/A}$  présentent une fixation à l'ARNt en retard sur gel comparable au sous-complexe sauvage  $P^H K$  (non montré).

Nous avons alors voulu savoir si ces mutations affectent l'activité de biosynthèse de t<sup>6</sup>A. De manière intéressante, la combinaison  $P^H K_{D/A} + BC^H$ , où le site de liaison au nucléotide de K est muté, ne présente plus aucune activité t<sup>6</sup>A. La combinaison  $P^H K_{H/A} + BC^H$ , où le site de coordination du fer de K est muté, présente très forte diminution de l'activité t<sup>6</sup>A (90%). (FIGURE II-11). Dans les deux cas, l'addition de protéine K sauvage aux mélanges permet de restaurer partiellement l'activité t<sup>6</sup>A.

Ces résultats indiquent que le site de liaison à un nucléotide par K est indispensable à la biosynthèse de t<sup>6</sup>A. De plus, la présence de l'atome de Fer semble jouer un rôle essentiel à cette activité, bien qu'une activité t<sup>6</sup>A résiduelle puisse être obtenue en son absence, indiquant qu'il n'est pas strictement nécessaire.



## Discussion

### Stabilité des sous-unités et des sous complexes de KEOPS :

Alors que les protéines P et B sont totalement insolubles après la première étape de purification, la présence de leur partenaire respectif permet de les maintenir en partie sous forme soluble, indiquant un rôle stabilisateur des partenaires les uns envers les autres. Il est à noter que tous les précipités observés lors de la purification des sous-complexes contiennent une quantité équimolaire de chaque protéine, indiquant donc que les interactions entre partenaires sont maintenues. De même, la coloration rose observée au niveau des précipités PK, KB et KBC indique que le site actif de K est toujours dans une conformation native. Ces observations suggèrent donc que cette précipitation ne provient pas d'une dénaturation de la structure tertiaire de chaque sous-unité, mais plutôt d'un phénomène d'agrégation dû à certaines interactions aspécifiques entre sous-complexes correctement pré-assemblés. Ce phénomène est probablement dû à des conditions *in vitro* éloignées des conditions *in vivo*. De manière intéressante, nous avons observé que la formation de tels précipités semble favorisée par les basses températures. Il est connu que certaines protéines, sous l'effet du froid, voient certains résidus hydrophobes exposés au solvant (DIAS ET AL., 2010), ce qui pourrait engendrer des interactions inter-protéines non spécifiques. Les protéines étudiées ici proviennent de l'archée hyperthermophile *P.abysyi*, dont l'optimum de croissance se situe autour des 90°C. Ainsi, la sensibilité de ces protéines aux basses températures observées ici traduirait en réalité une adaptation aux fortes températures.

### Rôle des différentes sous-unités :

Les différents résultats obtenus à l'aide des sous complexes nous permettent dans un premier temps d'établir un rôle général pour chacune des sous-unités du complexe KEOPS, que nous discuterons par la suite.

- La protéine Kae1 est impliquée dans la fixation de l'ARNt et au transfert du groupement TC sur cette molécule.
- La protéine Bud32 est responsable de l'hydrolyse de l'ATP en ADP+Pi. Cette activité est nécessaire à la synthèse de t<sup>6</sup>A, et est modulée par la protéine Kae1.

- La protéine Cgi121 permet de former un complexe KEOPS stable, et permet à l'activité ATPase de Bud32 d'être stimulée par la présence d'ARNt. Elle n'est pas strictement nécessaire mais augmente l'activité de biosynthèse de t<sup>6</sup>A.
- La protéine Pcc1 est impliquée dans la fixation de l'ARNt, la dimérisation *in vitro* du complexe et elle permet également à Bud32 de répondre à l'ARNt aux côtés de Cgi121. Contrairement à cette dernière, elle est strictement nécessaire pour la réaction t<sup>6</sup>A.

## Kae1 est lieu de transfert de l'AMP-TC et le centre de fixation de l'ARNt

La résolution de la structure de la protéine Kae1 de *P.abysyi* a montré qu'une molécule d'AMPPNP, analogue non hydrolysable de l'ATP, se retrouve au niveau du site actif (HECKER ET AL. 2007). De plus, chez *S.cerevisiae*, la mutation du résidu conservé impliqué dans la fixation de ce nucléotide affecte fortement la croissance cellulaire (MAO ET AL., 2008). Il a ainsi été suggéré que la fonction cellulaire de Kae1 soit liée à une activité ATPase.

Or, nos résultats indiquent Kae1 n'est pas responsable de l'activité ATPase du complexe KEOPS. En revanche, la mutation du site de fixation au nucléotide provoque une perte totale de l'activité de synthèse de t<sup>6</sup>A. L'étude de la structure des protéines TobZ et HypF (PARTHIER ET AL., 2012, PETKUN ET AL. 2011) suggère par analogie que Kae1 est le centre catalytique responsable de la fixation de l'intermédiaire TC-AMP et du transfert du groupement TC sur l'ARNt lors de la synthèse de t<sup>6</sup>A. Nos résultats supportent donc ce modèle.

De même, la mutation chez *S.cerevisiae* du site de coordination du Fer de Kae1 affecte fortement la croissance (MAO ET AL., 2008). La perte importante d'activité de biosynthèse de t<sup>6</sup>A chez notre mutant confirme l'importance de ce site de coordination, et indique que le Fer doit intervenir directement dans le mécanisme catalytique. La structure de Kae1 de *P.abysyi* montre que ce métal coordonne le phosphate γ terminal de la molécule d'AMPPNP, qui n'est pas présent sur l'intermédiaire TC-AMP. On peut donc suggérer par analogie que le Fer coordonne en réalité le groupement thréonine de l'intermédiaire.

De manière comparable, la protéine YgjD, orthologue de Kae1 dans la machinerie bactérienne de biosynthèse de t<sup>6</sup>A, doit jouer également un rôle similaire de fixation du TC-

AMP et du transfert du TC sur l'ARNt. De manière intéressante, il a été montré récemment que cette protéine présente un atome de Zinc au lieu d'un atome de Fer dans son site actif (NICHOLS ET AL., 2013). La présence d'un métal divalent semble donc être une caractéristique conservée pour ce domaine catalytique, mais sa nature ne semble pas être déterminante. La caractérisation du rôle exact de ce métal dans la fixation/catalyse de transfert de l'intermédiaire permettrait de mieux comprendre les différences existantes entre les machineries de biosynthèse de t<sup>6</sup>A des Archées et des Bactéries.

Ce rôle de transfert du groupement TC de Kae1 implique que l'ARNt soit fixé directement par cette protéine. Il a été montré précédemment que la protéine Kae1 peut se fixer de manière coopérative et aspécifique sur de l'ADN ou de l'ARN double brin et simple brin *in vitro* (HECKER ET AL., 2007, ET DONNEES NON PUBLIEES). Nos résultats indiquent que cette protéine présente également un comportement comparable pour les ARNt, en accord avec son rôle prévu dans la biosynthèse de t<sup>6</sup>A. L'étude *in cellulo* de la machinerie homologue chez le Xénope (MORIN ET AL., 1998) indique que les déterminants majeurs pour la reconnaissance des ARNt substrats se situent au niveau de la boucle anticodon, plus précisément aux positions adjacentes à la base modifiée 37, et correspondant à la un U en position 36 et un A en position 38. Ces positions doivent donc se retrouver à proximité immédiate du site actif de Kae1, ce qui suggère que cette protéine est le principal élément responsable de la discrimination des ARNt substrats et non substrats. Il serait alors intéressant de savoir quels sont les résidus de Kae1 qui interagissent avec l'ARNt (seule ou en présence de ses autres partenaires), par des techniques de *cross-linking*, afin de comprendre le mécanisme de sélection des ARNt substrats. De manière similaire, l'étude des propriétés de fixation à l'ARNt par la protéine YgjD permettrait de comparer les modalités de reconnaissance des différentes machineries vis-à-vis de l'ARNt.

Cependant, la protéine Kae1 seule possède une affinité relativement faible pour les ARNt et présente un comportement de fixation du type « *tout ou rien* » uniquement à forte concentration pour les ARNt. Nous avons montré que la présence des autres sous-unités module cette fixation, en augmentant fortement l'affinité et en modifiant ce comportement, indiquant que Kae1 n'est pas le seul élément impliqué dans la reconnaissance de l'ARNt. MORIN ET AL. ont montré qu'une structure tridimensionnelle intacte de l'ARNt est l'autre déterminant majeur pour permettre la synthèse de t<sup>6</sup>A chez les Eucaryotes. On peut donc

suggérer que les partenaires de Kae1 sont impliqués dans la reconnaissance de ce paramètre.

### L'activité ATPase de Bud32 traduit un mécanisme d'adaptation induite de conformation qui module l'activité de Kae1 dans la biosynthèse de t<sup>6</sup>A

Nous avons montré que Bud32 est responsable de l'hydrolyse de l'ATP en ADP + Pi libre observée lors de la synthèse de t<sup>6</sup>A par le complexe KEOPS, et que cette activité nécessite la présence de Kae1. Il avait été montré précédemment que Kae1 inhibe l'activité d'autophosphorylation de Bud32 (HECKER ET AL, 2009). Nous avons pu préciser l'action de Kae1 en montrant qu'elle agit en réalité comme un régulateur qui favorise la libération de Pi libre par Bud32 en inhibant l'activité de phosphotransférase. La structure de la protéine MJ1130 de *Methanocaldococcus jannaschii*, correspondant à une fusion de Kae1 et Bud32, révèle que certains résidus de Kae1 interagissent directement au niveau du site actif de Bud32 (HECKER ET AL, 2008, MAO ET AL, 2008). Ces résidus sont donc potentiellement impliqués dans la régulation de l'activité de Bud32

Bud32 a été décrite initialement comme une Sérine/thréonine kinase (STOCCHETTO ET AL, 1997), mais la question de la réalité de son activité kinase restait jusqu'à présent une question ouverte. En effet, certains éléments fondamentaux caractéristiques des kinases d'Eucaryotes sont absents chez Bud32, notamment la boucle A impliquée dans la reconnaissance du substrat, et la lysine conservée dans le site actif responsable du phosphotransfert, ce qui la classait parmi les kinases atypiques (HECKER ET AL., 2008, FACCHIN ET AL., 2002 ). Récemment, il a été démontré que la protéine Rio2, une autre kinase atypique proche de Bud32 au niveau structural et manquant des mêmes éléments, agit comme une ATPase et non comme une kinase (FEIRERRA CERCA ET AL., 2012). Nos données indiquent clairement que c'est le cas également pour Bud32 qui joue le rôle d'une ATPase au sein du complexe KEOPS.

On peut cependant suggérer que, dans d'autres conditions et en l'absence de Kae1, l'activité kinase soit prédominante. Or, chez certaines Archées comme *M.jannaschii*, les deux protéines sont fusionnées: il semble donc peu probable que chez ces organismes, l'activité phosphotransférase de Bud32 ait une réalité fonctionnelle.

D'un point de vue évolutif, il a été suggéré que Bud32 et les Rio kinases représentent des versions ancestrales des kinases eucaryotes modernes (LEONARD ET AL., 1998). Ainsi, d'après nos résultats, nous pouvons donc suggérer que l'activité phosphotransférase de ces protéines a émergé à partir de l'activité ATPase d'origine, par l'addition de certains éléments et résidus absent chez Bud32 et les Rio.

Le rôle de cette activité ATPase dans le mécanisme de biosynthèse de  $t^6A$  est l'un des enjeux majeurs dans la compréhension de ce système. Nous avons montré que l'hydrolyse de l'ATP par Bud32 est indispensable à cette réaction. De plus, nous avons précédemment montré que cette activité ATPase est fortement stimulée par la présence spécifique d'ARNt (PERROCHIA ET AL., 2013). Nous avons montré ici que cette sensibilité à l'ARNt, permise par la présence de Pcc1 et Cgi121, n'est pas strictement nécessaire à l'activité  $t^6A$  mais la stimule fortement. L'hydrolyse de l'ATP pourrait alors traduire un mécanisme d'adaptation induit de conformation du complexe contrôlant la réaction de catalyse, l'ARNt constituant le déclencheur de ce mécanisme. Les données structurales montrent en effet que la queue C-terminale de Bud32 se projette au niveau du site actif de Kae1 (MAO ET AL., 2008). On peut donc suggérer que la fixation d'un ARNt au niveau de Kae1, et peut être de Bud32, active l'hydrolyse de l'ATP par Bud32, ce qui provoque un changement conformationnel de la protéine qui modifie en retour les propriétés structurales du site actif de Kae1, ou qui positionne correctement l'ARNt à ce niveau, permettant à l'étape terminale du transfert du TC sur l'ARNt.

Cette hydrolyse pourrait constituer également un mécanisme de contrôle dans la discrimination du substrat ARNt à modifier. L'activité ATPase de Bud32 semble pouvoir être stimulée par tout type d'ARNt (PERROCHIA ET AL., 2013), mais la détermination précise des paramètres cinétiques d'hydrolyse de l'ATP en présence de différentes espèces d'ARNt permettrait de déterminer si des différences sont détectables entre des ARNt substrats et non substrats, et de savoir si cette activité joue un rôle dans la discrimination du substrat.

Enfin, cette activité pourrait intervenir dans l'étape de relargage de l'ARNt après sa modification, à l'image de Rio2. Cette protéine se fixe sur l'ARN ribosomal de la particule 40S en cours de maturation, et il a été montré que son activité ATPase est nécessaire pour

relarguer l'ARN à la fin de la maturation (FEIRERRA-CERCA ET AL., 2012). La détermination des constantes de dissociation entre KEOPS et l'ARNt, en présence ou en l'absence d'ATP, est donc une autre voie à explorer pour déterminer si Bud32 est impliquée dans un mécanisme similaire.

De manière intéressante, la machinerie bactérienne présente elle aussi une activité d'hydrolyse de l'ATP en ADP+Pi, réalisée par les protéines DEZ réunies. Les données disponibles ne permettent pas de savoir laquelle de ces protéine est responsable de cette activité. Pourtant, il est à noter qu'YjeE possède tous les éléments structuraux d'une ATPase à P-loop. De plus, il a été suggéré qu'YjeE agisse comme un interrupteur moléculaire, c'est-à-dire que l'hydrolyse de l'ATP provoque des changements conformationnels pouvant être transmis à ses partenaires (TEPLYAKOV ET AL., 2002). Ces caractéristiques rappellent de manière étonnante nos observations, où Bud32 semble jouer le rôle d'un interrupteur moléculaire. Cependant, l'hydrolyse de l'ATP par DEZ n'est pas dépendante de la présence d'ARNt (DEUTSCH ET AL., 2012), contrairement à KEOPS. De plus, il a été montré que le transfert du TC sur l'ARNt par les protéines DEZ peut être réalisé sans l'addition d'ATP (LAUHON ET AL., 2013), alors que l'hydrolyse de l'ATP est indispensable à la synthèse de t6A dans le cas de KEOPS. Ces données suggèrent donc que l'importance et le rôle de l'hydrolyse de l'ATP ne sont pas nécessairement identiques entre les deux systèmes.

Récemment, il a en effet été montré par microcalorimétrie qu'YgjD et YeaZ forment spontanément un hétérodimère en solution, alors que l'interaction entre YjeE et le dimère YgjD/YeaZ est soumise à une dynamique complexe, dépendant strictement de la présence d'ATP ou d'ADP (NICHOLS ET AL., 2013). Ainsi, l'hydrolyse de l'ATP dans le système bactérien pourrait être nécessaire à la stabilisation de ce système dynamique à trois composants. L'analyse détaillée du système bactérien, à l'image du travail effectué ici, et la détermination de la protéine responsable de l'hydrolyse de l'ATP permettrait de comprendre les différences qui existent entre les machineries.



## La protéine Cgi121, chaperone moléculaire impliquée dans la stabilité de KEOPS ?

Nos essais de reconstitution par filtration sur gel montrent de manière surprenante que le mélange PK+KB ne forme pas un complexe ternaire PKB isolable dans les conditions standards, alors que ce mélange possède une activité de biosynthèse de  $t^6A$ . Une incubation des sous-complexes PK+KB à une température de 50°C, en présence d'ARNt, d'ATP et des différents cofacteurs, ne permet pas non plus d'isoler un complexe PKB en filtration sur gel (données non montrées). Nous avons également tenté d'obtenir sans succès par plusieurs stratégies de co-expression le complexe PKB. Ces données suggèrent que l'interaction simultanée des trois protéines P, K et B est hautement dynamique et/ou instable. Il serait intéressant d'étudier la dynamique des interactions P, K et B par microcalorimétrie ou des techniques plus sensibles comme la SPR (*surface plasmonic resonance*, ou résonance plasmonique de surface).

Alors que le complexe ternaire PKB n'a pas pu être obtenu, le complexe PKBC est en revanche très stable en solution. De plus, l'ajout de C à PK+KB permet d'obtenir un complexe PKBC stable en gel filtration. La protéine C apparaît donc jouer un rôle de chaperonne pour l'ensemble du complexe en stabilisant ses interactions intermoléculaires. De plus, l'ajout de C au mélange PK+KB augmente la production de  $t^6A$ . Ainsi, l'effet stabilisateur de C, via B, permettrait d'augmenter l'efficacité de la catalyse de  $t^6A$ . Ainsi, il se pourrait que les Bactéries et les Archées Eucaryotes aient développé deux stratégies différentes visant à stabiliser l'interaction entre les composants de la machinerie de synthèse de  $t^6A$  : la présence d'un facteur protéique supplémentaire chez les Archées et les Eucaryotes, et la fixation de nucléotide chez les Bactéries.

Nous avons également montré que C est indispensable pour obtenir une stimulation de l'ATPase de Bud32 par les ARNt, alors qu'elle n'est pas impliquée directement dans la fixation de l'ARNt. La protéine C pourrait donc influencer la manière dont KEOPS reconnaît l'ARNt en modifiant la conformation du site actif de Bud32, afin de permettre à l'activité ATPase d'être sensible à la présence d'ARNt.

Ces hypothèses sur le rôle structural de Cgi121 sur Bud32 sont soutenues par les données de Mao *et al.* qui montrent que la présence de C au niveau du complexe KB provoque un changement conformationnel notable dans la structure de B. (MAO ET AL., 2008).

Enfin, nous avons observé un épaulement à 305nm dans le spectre d'absorbance de la protéine Cgi121 purifiée. Un épaulement similaire mais non identique a été retrouvé dans le spectre du complexe KEOPS entier (purifié en 200mM de NaCl), centré sur 340nm. Ces absorptions particulières pourraient être dues à la présence d'un cofacteur particulier de nature inconnue liée à cette protéine. Cependant, nous n'avons pas pu l'obtenir de manière systématique chez la protéine C, et les sous-complexes BC et KBC ne présentent pas cet épaulement. Ainsi, la présence d'un potentiel cofacteur pourrait dépendre de certaines conditions que nous n'avons pas pu déterminer, et qui ne semble pas indispensable à la synthèse de t<sup>6</sup>A.

### La protéine Pcc1, facteur de dimérisation et de fixation aux ARNt

Alors que C semble être impliquée dans la reconnaissance des ARNt via une action sur B lors de l'hydrolyse de l'ATP, la protéine P semble en revanche être directement liée à la fixation à l'ARNt. Nous avons mis en évidence que le complexe PK présente une forte affinité pour les ARNt en retard sur gel, meilleure que celle de la protéine K seule et comparable au complexe PKBC. De plus, la présence de P est strictement nécessaire à la biosynthèse de t<sup>6</sup>A et à une stimulation de l'hydrolyse de l'ATP par l'ARNt.

La protéine P issue de *Pyrococcus furiosus* a été décrite précédemment sous forme dimérique en solution et il a alors été proposé que le complexe KEOPS entier se dimérise via P *in vivo* (MAO ET AL., 2008). Chez l'Homme, il a également été montré que l'orthologue de P est responsable de la dimérisation du complexe PK (COSTESSI ET AL., 2012)

Nos données semblent confirmer ce modèle, car le complexe P<sup>H</sup>KBC et le sous-complexe P<sup>H</sup>K présentent une masse moléculaire apparente compatible avec celle d'un dimère ; alors que tous les sous-complexes purifiés en l'absence de P présentent un état monomérique. La mutation des résidus impliqués dans l'interface de dimérisation de P chez *S.cerevisiae* provoque un effet similaire à une délétion de P, ce qui indique que l'état dimérique a une importance fonctionnelle et est nécessaire à la fonction du complexe KEOPS (MAO ET AL.,

2008). De manière intéressante, de nombreuses amino-acyl ARNt transférases de classe II sont sous forme dimériques, il a été montré que cette dimérisation est nécessaire pour la fonction biologique de certaines de ces enzymes, qui implique notamment une fixation et une discrimination de l'ARNt substrat (ERIANI ET AL., 1993). L'état dimérique engendré par P pourrait être impliquée dans un processus de fixation/reconnaissance de l'ARNt. Il serait donc intéressant de produire des sous-complexes PK mutés au niveau des résidus responsables de la dimérisation de P, afin de mesurer l'effet de ces mutations sur la fixation aux ARNt.

Les données structurales soutiennent également le rôle de P dans la fixation directe de l'ARNt que nous proposons. En effet, cette protéine présente un repliement très similaire à celui du domaine KH (MAO ET AL., 2008), domaine structurale inclus dans de très nombreuses protéines et qui a pour rôle la fixation d'ARN ou ADN simple brin (VALVERDE ET AL., 2008). P ne possède pas le motif spécifique GXXG présent dans l'interface de fixation des domaines KH, mais il a été montré récemment que des domaines KH possédant un tel motif altéré ou absent pouvaient être fonctionnels (BRYKAILO ET AL., 2007). Ainsi, P serait une protéine à domaine KH atypique, et pourrait même représenter une version ancestrale du domaine KH en raison de sa présence à la fois chez les Archées et les Eucaryotes. Son rôle au sein du complexe KEOPS serait donc de participer à la fixation des ARNt aux côté de la protéine K, et éventuellement jouer un rôle dans les modalités de reconnaissance du substrat. De manière intéressante, seuls 4 nucléotides peuvent être accommodés au sein d'un repliement KH classique, et interagissent de manière séquence-spécifique avec les résidus de la protéine. Or, au niveau de l'ARNt, la boucle anticodon est une région simple brin de 7 nucléotides déterminant dans la discrimination du substrat chez les Eucaryotes. Il serait donc intéressant de déterminer si la boucle anticodon ou une autre région simple brin de l'ARNt est liée à P, par étude structurale de complexe PK ou par expérience de *cross-linking*.

## La fixation *in vivo* aux ARNt nécessite l'ensemble du complexe KEOPS et permettrait d'étudier des modalités de reconnaissance

Bien que nous ayons montré que P et K soient les principaux éléments responsables de la fixation de l'ARNt, il apparaît que le complexe KEOPS complet présente des propriétés de fixation qui diffèrent de ses sous-complexes. Cette observation est soutenue par le fait que complexe KEOPS co-purifie avec certains ARNt issus de la souche d'expression, alors qu'aucun des sous-complexes et sous-unités purifiés (P, K, B, C, PK, KB, BC, KBC) ne montre cette association. La détermination par Northern Blotting de la nature des ARNt liés au complexe KEOPS dans ces conditions pourrait fournir des renseignements sur la spécificité et les modalités de reconnaissance envers les ARNt.

## Comparaison des données biochimiques avec les données génétiques disponibles sur le rôle des sous-unités dans la biosynthèse de t<sup>6</sup>A :

Au bilan, il apparaît d'après nos données que les protéines P, K, B sont strictement requises pour la synthèse de t<sup>6</sup>A, mais que l'ensemble des 4 sous-unités du complexe KEOPS jouent un rôle et sont nécessaires à une activité optimale. Cette conclusion est en accord avec les observations réalisées chez des mutants de *S.cerevisiae* (SRINIVASAN ET AL., 2011), où la délétion de P, K et B diminue la quantité de t<sup>6</sup>A *in vivo*, alors que la délétion de C ne provoque pas de changement détectable. La protéine C semble donc bien jouer le rôle d'un facteur accessoire nécessaire à la régulation de l'activité du complexe.

En revanche, les mutants de délétion des gènes codant pour les sous-unités de KEOPS chez l'Archée *Haloferax volcanii* ne fournissent pas les mêmes résultats (NAOR ET AL, 2012): alors que la délétion de K, B et C est létale, un mutant délété pour P reste viable, et présente une diminution relativement faible de la quantité de t<sup>6</sup>A *in vivo* (*chiffre*). Cette observation pourrait suggérer que dans la cellule, le sous-complexe KB(C) peut catalyser t<sup>6</sup>A sans recourir à P, peut-être grâce à certaines conditions cellulaires ou facteurs qui sont absentes dans nos essais *in vitro*.

Enfin, il n'est pas exclu que les différents membres de KEOPS soient également impliqués dans d'autres voies métaboliques. Nous avons démontré que chacun d'entre eux joue un rôle direct dans la biosynthèse de t<sup>6</sup>A. Leur mutation provoque de très nombreux phénotypes, dont certains, voire tous, peuvent probablement être expliqués par des problèmes traductionnels liés à l'absence de t<sup>6</sup>A. La détermination exacte des conséquences *in vivo* de l'absence de t<sup>6</sup>A est donc un enjeu majeur afin de valider cette hypothèse, et permettra peut-être d'ouvrir de nouvelles pistes. En effet, si certains phénotypes liés à l'absence de KEOPS ne peuvent être expliqués par l'absence de t<sup>6</sup>A, ils pourront alors constituer des voies de recherche pour trouver d'autres rôles possibles du complexe dans la cellule.

## Conclusion

A l'issue de cette étude, nous pouvons proposer un modèle de fonctionnement plus détaillé de la machinerie de synthèse de t<sup>6</sup>A chez les Archées.

La protéine K active la protéine B et la maintient sous une conformation lui imposant d'hydrolyser l'ATP en ADP+Pi libre, en inhibant son activité phosphotransférase. D'autre part, la protéine C modifie la conformation de B pour lui permettre de répondre à la présence d'ARNt, et/ou pour stabiliser le complexe PKB. La protéine P, aux côtés de la protéine K, fixe alors un ARNt non modifié. La présence de cet ARNt déclenche l'hydrolyse de l'ATP en ADP+Pi par B. L'énergie de cette hydrolyse engendre un changement conformationnel du complexe, qui permet de contrôler l'ARNt fixé ou de le positionner correctement au niveau de site actif de K. La protéine K, ayant fixé l'intermédiaire TC-AMP produit par Sua5, transfère alors le groupement TC sur l'ARNt. L'ARNt modifié est alors relâché.

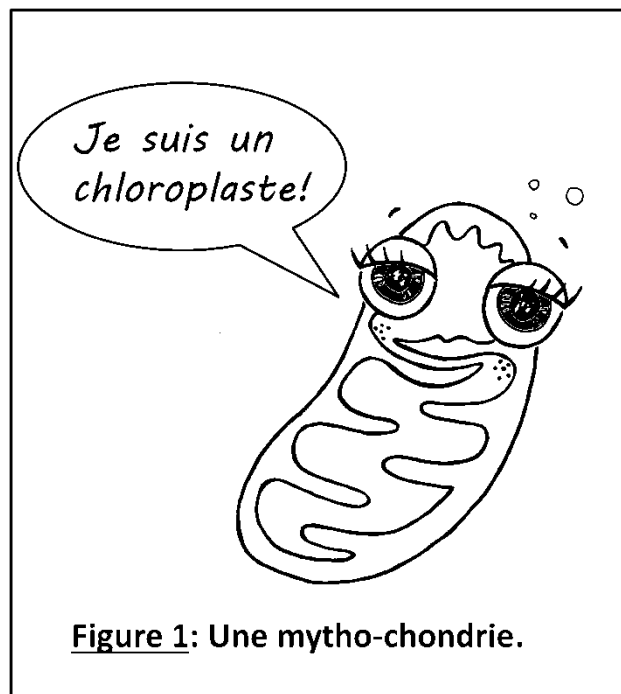
Dans un modèle alternatif, l'hydrolyse de l'ATP ne survient qu'après fixation et modification de l'ARNt, pour relarguer cette molécule à l'issue du transfert.

Les résultats présentés ici démontrent que l'étude *in vitro* des sous-complexes et des sous-unités de KEOPS constitue un outil approprié et performant permettant une meilleure compréhension de la voie de biosynthèse de t<sup>6</sup>A. Nous avons mis en évidence que le complexe KEOPS est une machinerie sophistiquée, régie par des interactions réciproques entre ses différents composants qui lui permettent de remplir de manière efficace sa

fonction. Certaines questions restent ouvertes, notamment au niveau des modalités de reconnaissance des substrats ARNt et du rôle exacte de l'hydrolyse de l'ATP par Bud32. Les différents résultats que nous avons obtenus permettent ainsi d'orienter les futurs travaux de caractérisation, afin de répondre à ces questions. De plus, par comparaison avec la machinerie analogue caractérisée chez les Bactéries, cette étude ouvre la voie à la compréhension des différentes stratégies mises en place par les organismes pour effectuer cette réaction complexe. Enfin, le modèle établi dans cette étude permet également de servir de base au fonctionnement du complexe KEOPS chez les Eucaryotes.

Chez ces organismes, deux machineries de synthèse de  $t^6A$  semblent co-exister : d'une part, un système semblable à celui des Archées, formé de la protéine Sua5 et du complexe KEOPS ; qui modifie les ARNt cytoplasmiques ; et d'autre part, un système mitochondrial, dont ferait partie l'orthologue de Kae1/YgjD : Qri7. Notre étude et les études similaires sur les protéines bactériennes montrent que la présence de partenaires spécifiques à Kae1/YgjD est indispensable à la catalyse de formation de  $t^6A$ . Il est donc intéressant de comprendre quelle est la nature et le mode de fonctionnement de la machinerie mitochondriale. Cette étude fera l'objet d'une troisième et dernière partie.

# PARTIE III : ETUDE DE LA MACHINERIE MITOCHONDRIALE DE BIOSYNTHESE DE t<sup>6</sup>A







## Introduction

Les Eucaryotes présentent une organisation cellulaire complexe, caractérisée par la présence de différents compartiments : noyau, réticulum endoplasmique, appareil de Golgi et système vésiculaire associé, organites semi-autonomes. La mitochondrie représente un compartiment particulier, de part sa distribution et son origine. Presque toutes les lignées Eucaryotes possèdent cet organite, qui joue un rôle essentiel dans le métabolisme oxydatif et la respiration (les rares cas d'absence de mitochondrie s'expliquent en réalité par une perte secondaire, comme chez les Microsporidies, [GERMOT ET AL, 1999](#)). L'origine de cet organite a été comprise il y a maintenant plusieurs décennies : il s'agit d'un endosymbiote descendant d'une  $\alpha$ -protéobactérie. Cette théorie endosymbiotique, initialement formulée par Lynn Margulis ([L.SAGAN, 1967, ON THE ORIGIN OF MITOSING CELL](#)), a été confirmée par de nombreuses approches de phylogénie moléculaire, notamment par le travail de Carl Woese sur l'ARN 16S, démontrant une proximité entre l'ARNr bactérien et l'ARNr mitochondrial ([CR WOESE, ENDOSYMBIONTS AND MITOCHONDRIAL ORIGINS, 1977](#)). D'autres travaux ont montré que l'évènement d'endosymbiose ne s'est produit qu'une seule fois au cours de l'évolution, et bien avant la divergence des lignées Eucaryotes actuelles ([DESMOND ET AL, 2011](#)). Dans les cellules modernes, les mitochondries reflètent encore leur origine bactérienne par certains aspects : elles possèdent un génome circulaire codant un certain nombre de protéines et d'ARN, une machinerie de traduction fonctionnelle de type bactérien, et se divisent de manière semi-autonome au sein du cytoplasme de leur hôte.

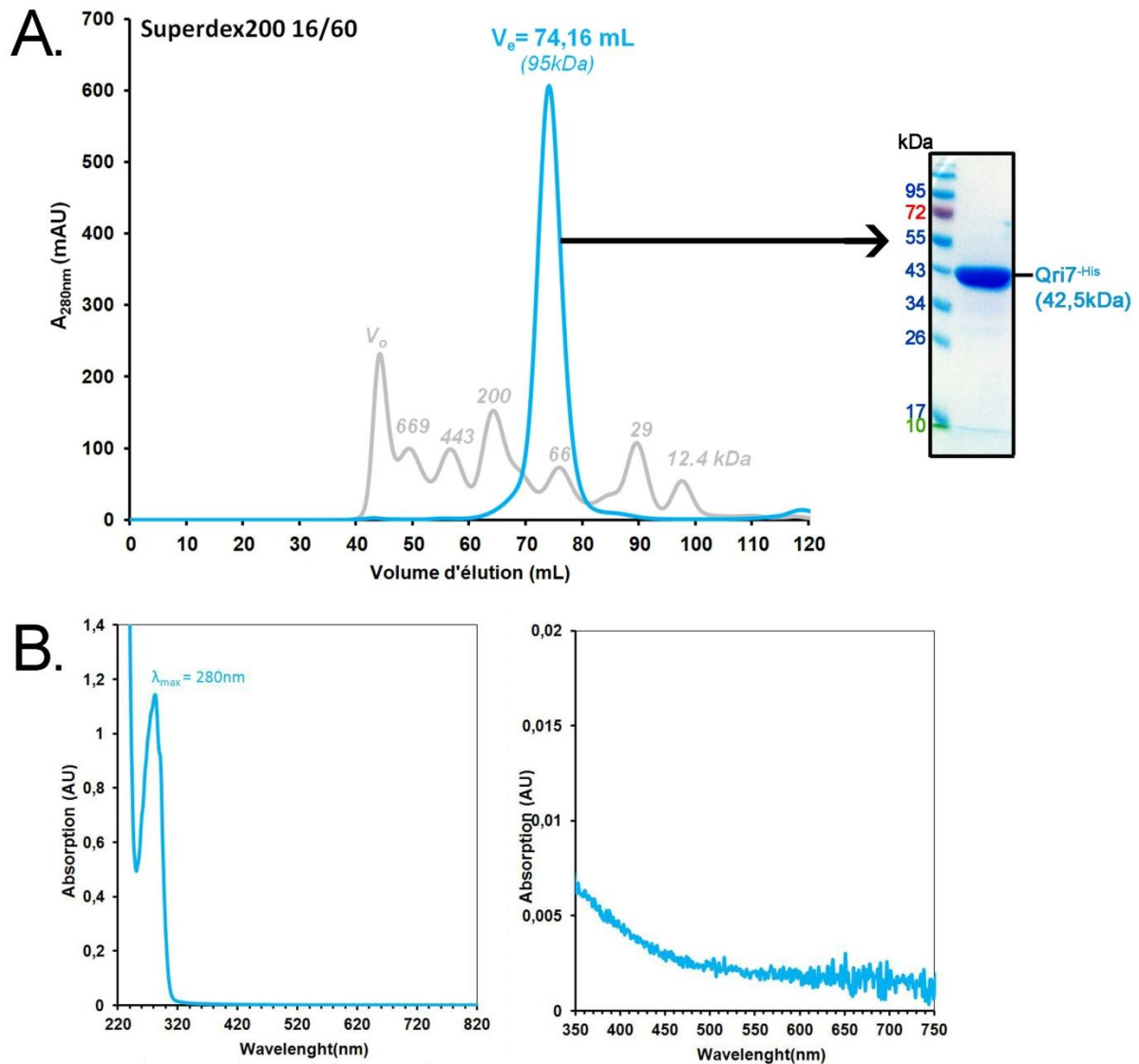
Chez la Levure *Saccharomyces cerevisiae*, organisme Eucaryote modèle, le génome mitochondriale a une taille de 85,8 kb, et encode seulement 30 protéines ([ANDERSSON ET AL, 2002](#)). Pourtant, plus de 800 protéines ont été expérimentalement assignées à cet organite ([SICKMANN ET AL, 2003](#)). Une analyse phylogénétique menée sur 400 de ces protéines a montré que 50% d'entre elles ont une origine Eucaryote, et 50% ont clairement une origine bactérienne et sont des orthologues de protéines présentes chez les  $\alpha$ -protéobactéries. ([KARLBERG ET AL, 2000](#)). Cette distribution reflète une caractéristique générale de l'évolution mitochondriale : la bactérie d'origine a perdu un certain nombre de protéines, remplacées par des équivalents eucaryotes, et a transféré la majeure partie de ses gènes bactériens restants au sein du génome de la cellule hôte. Un mécanisme d'import s'est parallèlement mis en place, afin rediriger ces différentes protéines au sein de l'organite ([ANDERSSON ET AL,](#)



2002, KARLBERG ET AL., 2000). Généralement, les protéines destinées à la mitochondrie présentent un signal d'adressage identifiable qui est par la suite clivé.

Ainsi, la mitochondrie actuelle est une structure chimérique complexe, qui a conservé des traces de son origine bactérienne au niveau protéique mais qui possède aussi des protéines issues de son hôte. De nombreux processus à l'œuvre dans le cytoplasme ont donc leur contrepartie dans la matrice mitochondriale: par exemple, les mitochondries effectuent la transcription, puis la traduction des protéines codées dans leur génome, avec leur propres ribosomes et en utilisant une panoplie d'ARNt mitochondriaux fonctionnels. Chez *S. cerevisiae*, 24 ARNt sont codées dans le génome mitochondrial et suffisent à lire tous les codons. Ces ARNt mitochondriaux présentent des nucléosides modifiés, notamment t<sup>6</sup>A (MODOMICS). Ainsi, après avoir décrit la machinerie de biosynthèse de t<sup>6</sup>A chez les Bactéries (DEUTSCH ET AL., 2012), chez les Archées et chez les Eucaryotes (PERROCHIA ET AL., 2013, ET CE TRAVAIL DE THESE), il reste à comprendre comment t<sup>6</sup>A est synthétisée au niveau de la mitochondrie.

La plupart des Eucaryotes présentent deux homologues de la famille Kae1/YgjD. L'une de ces versions, appelée Kae1 en référence à son nom chez *S. cerevisiae*, fait partie du complexe KEOPS et est proche de la version Archéenne. Les analyses phylogénétiques ont montré que l'autre version, que nous appellerons Qri7 en référence à son nom chez *S. cerevisiae*, est en revanche proche de la protéine YgjD bactérienne, plus précisément des orthologues présents chez les  $\alpha$ -protéobactéries (HECKER ET AL., 2007, OBERTO ET AL., 2009). Cette proximité suggère une origine mitochondriale. Il a été effectivement montré expérimentalement chez plusieurs espèces (*S. cerevisiae*, *C.elegans*, *A.thaliana*, *H.sapiens*) que Qri7 est localisée dans la mitochondrie (OBERTO ET AL., 2009, HAUSSEHEL ET AL., 2009). La délétion ou l'inactivation de ce gène entraîne de nombreux défauts mitochondriaux chez *S. cerevisiae* et *C.elegans* : perte de l'ADN mitochondrial, défauts métaboliques, morphologie mitochondriale anormale (OBERTO ET AL., 2009). Chez *A.thaliana*, la délétion de Qri7 provoque un arrêt de la croissance de la plante (HAUSSEHEL ET AL., 2009). Ces données indiquent que Qri7 joue un rôle essentiel pour la mitochondrie. Récemment, il a été montré que Qri7 est impliquée, tout comme ses homologues Eucaryotes et Bactériens, dans la biosynthèse de t<sup>6</sup>A. Un mutant  $\Delta$ Kae1 de *S. cerevisiae* présente une quantité résiduelle de t<sup>6</sup>A, alors qu'un double mutant  $\Delta$ Kae1/ $\Delta$ Qri7 ne présente plus aucune trace de t<sup>6</sup>A (SRINIVASAN ET AL., 2011).



**FIGURE III-1: Purification de la protéine Qri7<sup>-His</sup> de *Saccharomyces cerevisiae***

La souche d'expression E.coli Rosetta2(DE3)pLysS a été transformée avec le plasmide pET28a::ScQri7-His, puis une colonie isolée à servi à inoculer 1 Litre de TB OverNightExpress +. Après une nuit de croissance, la culture a été centrifugée, et les cellules ont été lysées par sonication à 4°C dans un tampon TrisHCl 50mM, pH8.0, NaCl 500mM, Glycérol 10%,  $\beta$ -mercaptoéthanol 5mM en présence d'inhibiteurs de protéase. La fraction soluble du lysat a été purifiée par une première étape de chromatographie d'affinité sur NiNTA suivi d'une seconde étape de filtration sur gel sur une colonne SuperDex 200 16/60. La quantité de protéine en sortie de colonne a été suivi par mesure de l'absorbance à 280nm. Les fractions d'élution ont été analysées sur SDS PAGE, et celles contenant la protéine pure ont été réunies et concentrées. L'absorbance de la solution protéique a été mesurée entre 200nm et 1000nm.;

**A. Chromatogramme de filtration sur gel et analyse SDS-PAGE de la protéine Qri7<sup>-His</sup> purifiée.**

**B. Spectre d'absorption de la protéine purifiée.**

De plus, Qri7 peut remplacer fonctionnellement Kae1 pour la synthèse de t<sup>6</sup>A dans le cytoplasme chez *S. cerevisiae* si son signal d'adressage mitochondrial est délété (EL YACOUBI ET AL, 2011). Cette protéine semble donc bien faire partie d'une machinerie de biosynthèse de t<sup>6</sup>A présente dans la mitochondrie.

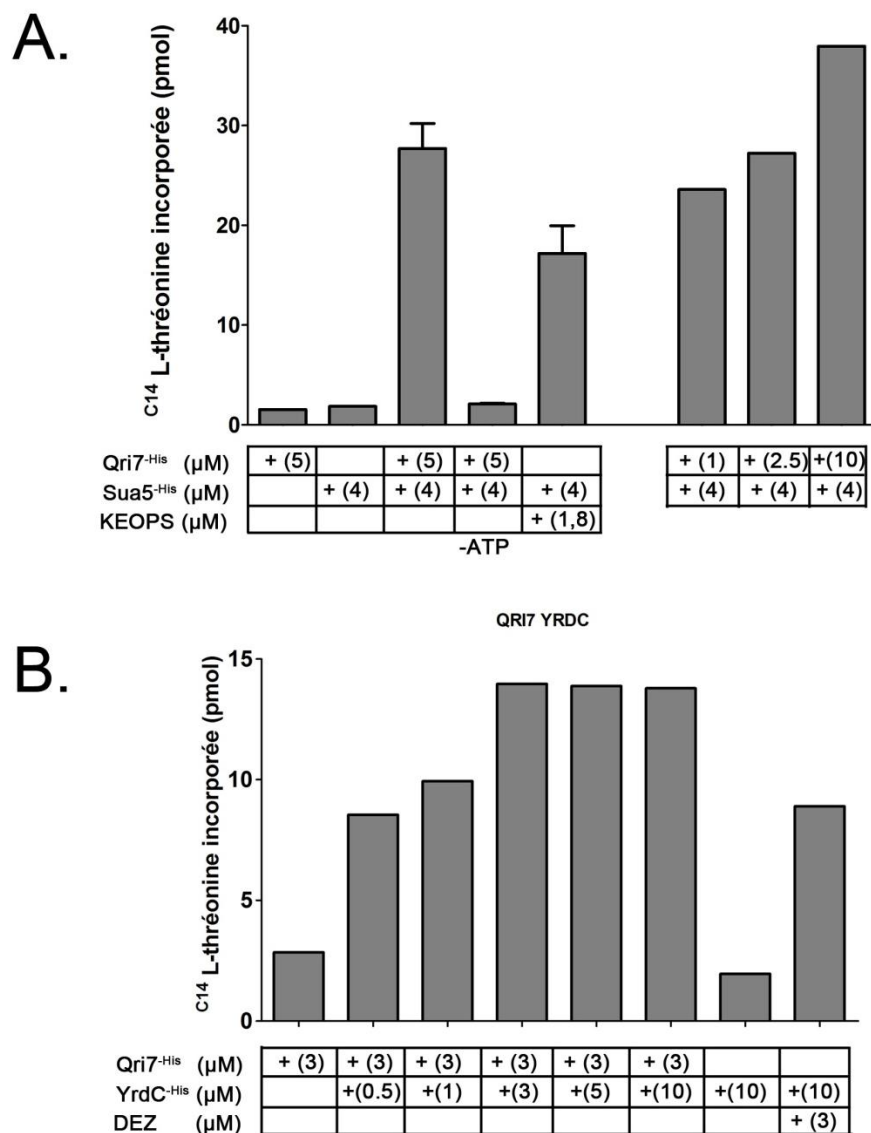
Cependant, la composition de cette machinerie mitochondriale est mystérieuse. En effet, aucun orthologue des protéines YeaZ et YjeE n'est détectable chez *S. cerevisiae*, et aucun des membres du complexe KEOPS n'a été localisé dans la mitochondrie par les études globales de protéomique. Ainsi, il semblerait que Qri7 puisse fonctionner sans l'aide des partenaires caractérisés au niveau des différentes machineries de biosynthèse de t<sup>6</sup>A. Afin de confirmer cette hypothèse, nous avons entrepris la purification de la protéine Qri7 et sa caractérisation biochimique.

## Résultats

### Purification de Qri7

Le gène codant la protéine Qri7 de *S. cerevisiae*, délété de la séquence codant les 27 premiers acides aminés correspondant au signal d'adressage mitochondrial, a été cloné dans un plasmide d'expression, fusionné à une étiquette hexahistidine en position 3' du gène, correspondant à l'extrémité C-terminale de la protéine. La protéine recombinante a été surexprimée dans une souche Rosetta2(DE3)pLys d'*Escherichia coli*. La protéine ainsi produite (notée Qri7<sup>(His)</sup>) d'une masse moléculaire de 42,5kDa, est retrouvée en très grande quantité dans la fraction soluble du lysat cellulaire et a été purifiée par une première étape de chromatographie d'affinité sur résine de nickel dans un tampon contenant 500mM de NaCl. Une seconde étape de filtration sur gel sur une colonne Superdex200 16/60 a permis de purifier jusqu'à homogénéité la protéine recombinante (FIGURE III-1A). Qri7<sup>(His)</sup> élue sous la forme d'un unique pic, à un volume de 74,16mL, correspondant à une masse moléculaire apparente d'environ 95 kDa, suggérant que cette protéine est sous forme dimérique en solution.

Le spectre d'absorption de la protéine purifiée ne présente aucun maxima à 500nm, suggérant que cette protéine ne contient pas d'atome de Fer au niveau de son site actif, contrairement à la protéine Kae1 de *Pyrococcus abyssi*. (FIGURE III-1B)



Note: les protéines YgjD, YjeE, YeaZ (DEZ) utilisées dans cette expériences proviennent d'alicots ayant subis un cycle décongelation/congélation.

### FIGURE III-2 : Activité t<sup>6</sup>A de la protéine Qri7<sup>-His</sup> avec Sua5<sup>-His</sup> et YrdC<sup>-His</sup>

Les protéines ont été incubées selon le protocole indiqué dans le Matériel et Méthodes, à 32°C pendant 40mn (A) ou à 37°C pendant 35mn (B).

Les ARNt utilisés sont : ARNt<sup>Lys(UUU)</sup> d'*Escherichia coli*, sur-exprimés in vivo (A) à une concentration de 4,8μM (240pmol), et ARNt<sup>Asn(AAC)</sup> de *Pyrococcus abyssi* sur-exprimés in vivo (B) à une concentration de 3,4 (170pmol)

La présence (+) et les concentrations (entre parenthèse, en μM) des protéines utilisées sont indiquées dans les tableaux sous chaque graphe.

**A. Activité de Qri7<sup>-His</sup> et Sua5<sup>-His</sup> de *S.cerevisiae*.**

**B. Activité de Qri7<sup>-His</sup> de *S.cerevisiae* et YrdC<sup>-His</sup> d'*Escherichia coli***

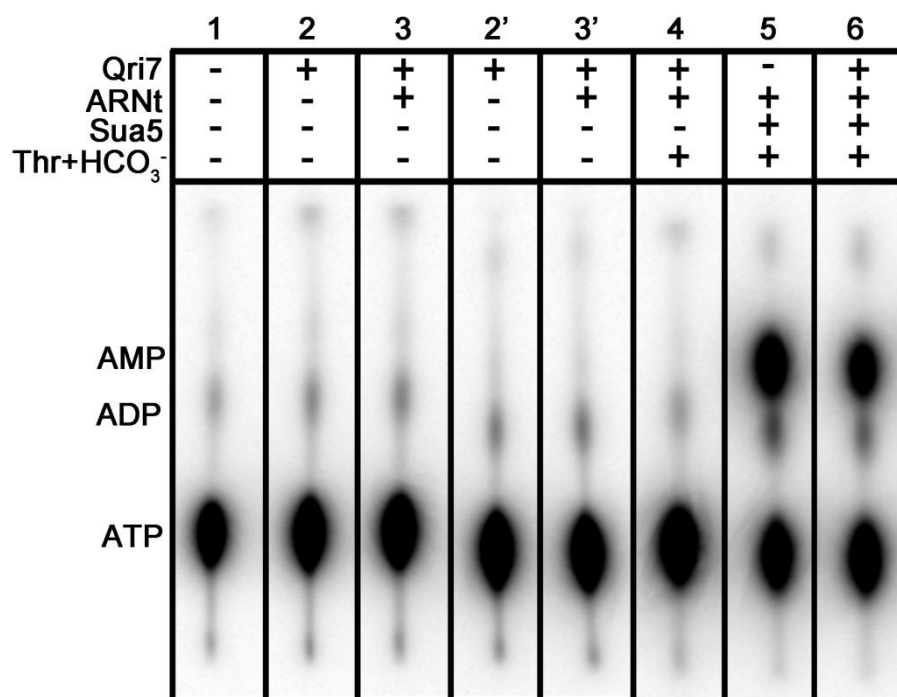
## Qri7 et Sua5 de *S. cerevisiae* catalysent ensemble la synthèse de t<sup>6</sup>A

Chez *S. cerevisiae*, un seul orthologue de Sua5 est détecté. Ainsi, cela suggère que la première étape de la biosynthèse de t<sup>6</sup>A, à savoir la production de TC-AMP, est réalisée par la protéine Sua5 au niveau mitochondrial, ou que l'intermédiaire est synthétisé dans le cytoplasme puis transféré dans la mitochondrie. Nous avons donc voulu savoir si Qri7 pouvait assurer la catalyse de la seconde étape de la biosynthèse de t<sup>6</sup>A *in vitro*.

L'incorporation de L-thréonine radiomarkée au Carbone 14 au niveau du matériel précipitable au TCA a été mesurée selon le protocole décrit précédemment, en présence d'ARNt<sup>Lys(UUU)</sup> d'*Escherichia coli*, de la protéine purifiée Qri7<sup>(His)</sup>, de la protéine purifiée Sua5<sup>(His)</sup> de *Saccharomyces cerevisiae*, de thréonine<sup>C14</sup>, de carbonate et d'ARNt<sup>Lys(UUU)</sup> d'*Escherichia coli*, après une incubation à 32°C pendant 40mn. En parallèle, nous avons suivi dans les mêmes conditions l'incorporation de thréonine par le mélange Sua5 et KEOPS de *S. cerevisiae* comme référence. (FIGURE III-2A)

La protéine Qri7 seule ne conduit à aucune incorporation de thréonine. En revanche, en présence de Sua5, on observe une incorporation importante de thréonine, qui dépend de la présence d'ATP. La quantité de <sup>C14</sup>L-thréonine incorporée par la combinaison Sua5+ Qri7 est similaire à celle incorporée par la combinaison Sua5+KEOPS dans les mêmes conditions, indiquant que l'efficacité des deux systèmes est comparable. Ayant montré précédemment que cette incorporation correspond à la formation de t<sup>6</sup>A au niveau des ARNt, nous avons pu calculer un rendement de modification d'ARNt d'environ 16% dans les conditions d'incubation testées, qui pourrait probablement être amélioré par une mise au point plus approfondie de ce système *in vitro*.

Ainsi, ces résultats indiquent que la protéine Qri7 est capable de catalyser le transfert du groupement TC sur l'ARNt.



**FIGURE III-3: Suivi de l'activité ATPase de Qri7<sup>-His</sup> et Sua5<sup>-His</sup> de *Saccharomyces cerevisiae***

Les différents composants indiqués (+/-) ont été incubés dans le tampon de réaction utilisé pour la réaction de synthèse de t<sup>6</sup>A in vitro, en présence d'ATP à une concentration de 100µM supplémenté de 1µCi d'ATP[αP32] dans un volume final de 20µL. Après 30mn d'incubation à 32°C, 1µL du mélange réactionnel a été déposé sur une plaque PEI-cellulose, et les produits d'hydrolyse de l'ATP ont été séparés par chromatographie dans un solvant KH<sub>2</sub>PO<sub>4</sub>, 0,5M, pH3.5.

La protéine Qri7<sup>-His</sup> se trouve à une concentration finale de 1µM dans les conditions 1, 2, 3, 4, et 6, et 10µM dans les conditions 2' et 3'.

La protéine Sua5<sup>-His</sup> est à une concentration de 10µM (5 et 6)

L'ARNt<sup>Lys(UUU)</sup> d'*E.coli* est à une concentration de 1µM (3,3',4,5,6), la L-thréonine à une concentration de 100µM et le NaHCO<sub>3</sub> à 5mM (4, 5, 6). Après migration et séchage, la plaque a été exposée sur un écran PhosphorScreen pendant une nuit.



## Qri7 et YrdC forment un système hétérologue fonctionnel de biosynthèse de t<sup>6</sup>A

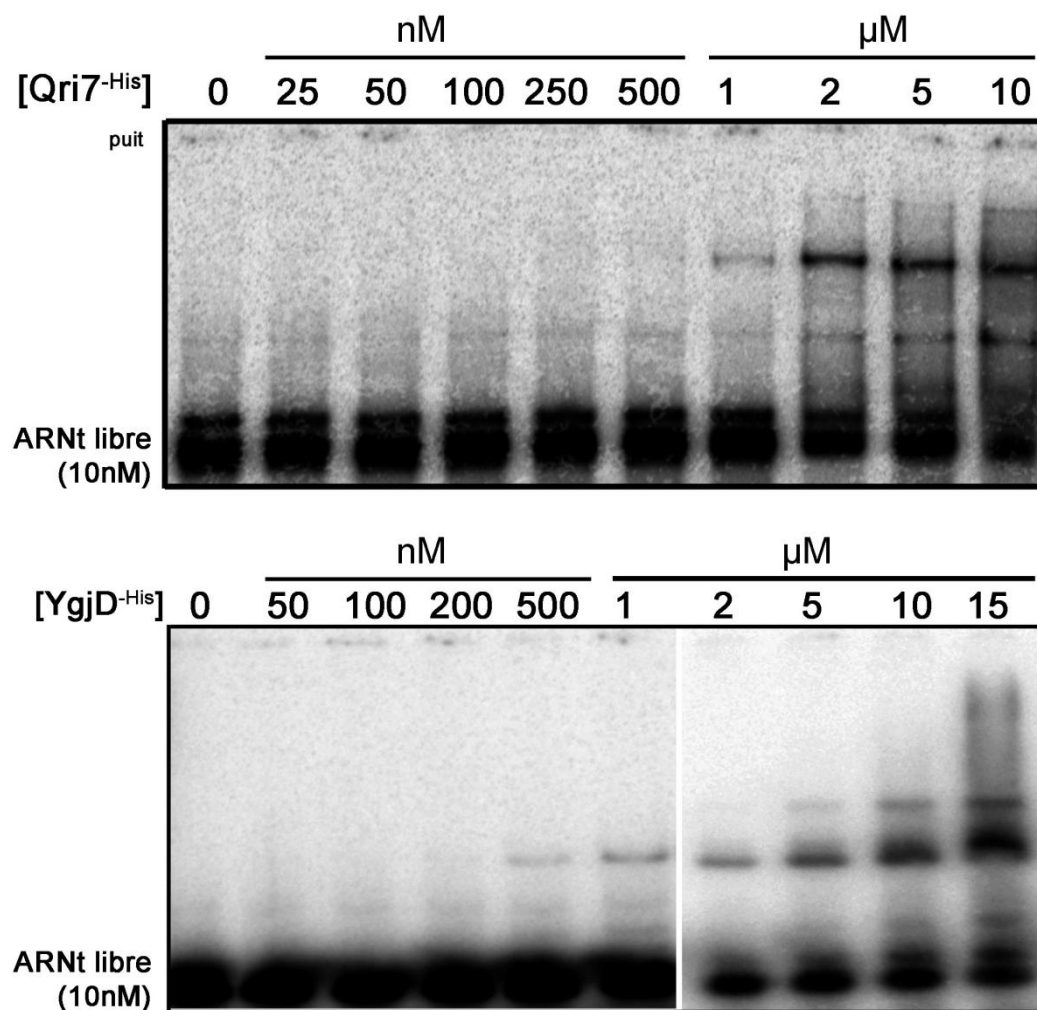
Différentes expériences de complémentations génétiques ont été effectuées chez *E. coli* afin de déterminer si Qri7 pouvait remplacer fonctionnellement la protéine YgjD, en utilisant des techniques d'inactivation différente et aboutissant à des résultats contradictoires (OBERTO ET AL., 2009, EL YACOUBI ET AL., 2011).

Nous avons donc voulu vérifier si la protéine YrdC d'*E. coli*, l'orthologue bactérien de Sua5, et la protéine Qri7 de *S. cerevisiae* peuvent constituer ensemble un système fonctionnel de biosynthèse de t<sup>6</sup>A. Nous avons incubé ces deux protéines avec les différents composants de la réaction (en utilisant ici des ARNt<sup>Asn(XXX)</sup> de *P. abyssi*) à 37°C, correspondant à la température de croissance d'*E. coli*. En référence, nous avons également suivi l'incorporation de <sup>14</sup>C-L-thréonine par le mélange YrdC+(YgjD+YjeE+YeaZ) (notées DEZ) incubé dans les mêmes conditions (FIGURE III-2B)

Le mélange Qri7+YrdC conduit à l'incorporation de <sup>14</sup>C-L-thréonine en quantité similaire à celui du mélange YrdC+DEZ, indiquant que Qri7 est fonctionnelle à 37°C pour la biosynthèse de t<sup>6</sup>A.

## Qri7 peut utiliser comme substrat des ARNt de différentes espèces :

Les deux expériences précédentes ont été réalisées avec des ARNt de séquences différentes : d'une part, l'ARNt<sup>Lys(UUU)</sup> de *Escherichia coli* (FIGURE III-2A), et l'ARNt<sup>Asn(XXX)</sup> de *Pyrococcus abyssi* (FIGURE III-2B) d'autre part. Cela indique que les modalités de reconnaissance de cette protéine reposent sur des déterminants présents chez ces deux types d'ARNt, et semble comparable aux machineries Eucaryotes et Archéennes décrites dans la partie I.



**FIGURE III-4:** Analyse de la fixation d'ARNt par Qri7<sup>-His</sup>, YgjD<sup>-His</sup> en retard sur gel de polyacrylamide en condition native.

Un mélange contenant une concentration variable de protéine, 10nM d'ARNt<sup>Lys(UUU)</sup> d'*E.coli* radiomarké, dans un tampon TrisHCl 50mM, pH8.0, MgCl<sub>2</sub> 5mM, DTT 1mM, NaCl 100mM, glycérol 10% à un volume final de 10μL a été incubé pendant 15mn à 4°C. 2,5μL Après ajout de 2,5μL de tampon de dépôt, les complexes nucléo-protéiques ont été séparé par électrophorèse sur un gel de polyacrylamide (19:1) 8%, TGE1X, 1mm d'épaisseur. Après 4 heures de migration à température ambiante sous une tension de 100V, le gel a été séché puis exposé une nuit sur un écran Phosphorscreen.

**A.** Retard sur gel avec la protéine Qri7<sup>-His</sup> de *Saccharomyces cerevisiae*

**B.** Retard sur gel avec la protéine YgjD<sup>-His</sup> d'*Escherichia coli* (deux expériences ont été réalisée avec des gammes de concentrations différentes)

## Qri7 n'utilise pas l'énergie d'hydrolyse de l'ATP pour la biosynthèse de t<sup>6</sup>A

Dans tous les systèmes de biosynthèse de t<sup>6</sup>A décrits précédemment (chez *E. coli*, *B. subtilis*, *S. cerevisiae* et *P. abyssi*), la deuxième étape de la réaction est associée à une hydrolyse de l'ATP en ADP + Pi. Nous avons alors suivi par chromatographie sur couche mince la consommation d'ATP par la protéine Qri7 dans les conditions de synthèse de t<sup>6</sup>A, en présence ou en l'absence des différents composants de la réaction (FIGURE III-3)

Dans tous les cas, aucune production d'ADP significative n'a été observée. Ainsi, la protéine Qri7 ne présente aucune activité ATPase associée à la biosynthèse de t<sup>6</sup>A.

## Qri7 présente une faible affinité pour les ARNt en retard sur gel, comparable à YgjD

La catalyse de formation de t<sup>6</sup>A au niveau des ARNt implique la fixation de l'ARNt au niveau du site de transfert de l'intermédiaire TC. Nous avons donc suivi la fixation de Qri7 sur des ARNt en retard sur gel. Une quantité croissante de protéine a été incubée avec une quantité fixe d'ARNt<sup>(Lys)UUU</sup> d'*E. coli* radiomarké, dans un tampon contenant du magnésium et une concentration de 100mM de NaCl. Après incubation à température ambiante pendant 15mn, les complexes nucléo-protéiques ont été séparés sur gel de polyacrylamide en conditions native. (FIGURE III-4A). On constate que la protéine Qri7 provoque un retard des ARNt détectable à partir de 1μM de protéine, correspondant à un ratio protéine/ARNt élevé (100), ce qui indique une affinité relativement faible dans ces conditions et pour cette espèce d'ARNt.

Nous avons également suivi la fixation de YgjD, l'orthologue de Qri7 chez *Escherichia coli*, sur l'ARNt dans les mêmes conditions. (FIGURE III-4B) Le profil obtenu est similaire à celui de Qri7.



## Discussion

### Qri7 et Sua5 constituent la machinerie de biosynthèse de t<sup>6</sup>A chez la mitochondrie

Précédemment, nous avons montré que les orthologues de la famille universelle Sua5/YrdC sont impliqués dans la première étape de la voie de biosynthèse de t<sup>6</sup>A. (PERROCHIA ET AL., 2013), à savoir la production d'un intermédiaire de réaction TC-AMP. Il est donc raisonnable de penser que la machinerie de biosynthèse de t<sup>6</sup>A chez la mitochondrie implique également un orthologue de cette famille de protéine. Or, chez la Levure comme chez les autres Eucaryotes, un seul représentant de cette famille est présent, et n'a pour le moment jamais été détectée au niveau de la mitochondrie. L'intermédiaire TC-AMP produit par cette protéine étant relativement instable (entre 3 à 16mn *in vitro* à 37°C, LAUHON ET AL., 2012), il semble donc peu probable qu'il puisse être produit au niveau cytoplasmique puis transféré dans la mitochondrie, où est localisée Qri7. Des résultats non publiés indiquent que la phase ouverte de lecture du gène Sua5 possède un codon d'initiation alternatif au codon d'initiation habituel, qui est à l'origine de la production d'une protéine possédant un signal d'adressage mitochondrial, et des expériences de localisation sub-cellulaire confirme la présence de Sua5 dans ce compartiment (Valérie de Crécy-lagard, communication personnelle). Ce double adressage est retrouvé chez de nombreuses protéines Eucaryotes, notamment l'enzyme de modification Mod5 responsable de la synthèse de i<sup>6</sup>G<sub>37</sub>, qui est dirigée dans le cytoplasme ou dans la mitochondrie chez la Levure selon l'utilisation d'un codon alternatif d'initiation. (BOGUTA ET AL., 1994), et l'enzyme Trm5, responsable de la synthèse de m<sup>1</sup>G<sub>37</sub> (LEE ET AL., 2007). Ces résultats et nos données *in vitro* indiquent donc que la machinerie de synthèse de t<sup>6</sup>A chez la mitochondrie est un système mixte à deux composants, constitué de la protéine Eucaryote Sua5 et de la protéine d'origine mitochondriale Qri7, et que la synthèse de t<sup>6</sup>A au niveau des ARNt mitochondriaux a lieu dans ce compartiment.

La plupart des Eucaryotes possèdent un orthologue mitochondrial de la protéine Qri7 (OBERTO ET AL., 2009). De manière intéressante, l'orthologue de Sua5 chez l'Homme, appelé IRIP, présente un signal mitochondrial identifiable (ANNOTATION ISSUE DE LA BANQUE DE DONNEE UNIPROT). Le même mécanisme de double adressage pourrait donc être à l'œuvre chez

l'Homme. Ainsi, le modèle de machinerie mitochondriale de biosynthèse de t<sup>6</sup>A que nous avons mis en évidence chez *S. cerevisiae* semble pouvoir être généralisable à l'ensemble des Eucaryotes.

La délétion ou l'inactivation de Qri7 chez *S. cerevisiae*, *C. elegans* et *A. thaliana* provoque différents phénotypes associés à de sévères défauts mitochondriaux. La mise en évidence direct du rôle de Qri7 dans la biosynthèse de t<sup>6</sup>A suggère que ces phénotypes sont probablement dus à l'absence de cette modification au niveau des ARNt mitochondriaux.

### La mitochondrie : un modèle simplifié pour l'étude du rôle *in vivo* de t<sup>6</sup>A.

A l'issue de cette étude et des études précédemment citées, les machineries de synthèse de t<sup>6</sup>A dans les différents domaines du vivant sont maintenant clairement identifiées. Tout comme pour la mitochondrie, la délétion des composants de ces machineries, chez différentes organismes, provoque une pléiotropie de phénotypes, allant de la mort de l'organisme à d'importantes perturbations de certains processus cellulaires. (OBERTO ET AL., 2009, HAUSSUEHL ET AL., 2009), probablement dus à l'absence de t<sup>6</sup>A.

Chez *S. cerevisiae*, il a en effet été montré que l'absence de t<sup>6</sup>A engendrée par la délétion de Kae1 et Sua5 provoque des initiations de traduction alternatives et une augmentation de la fréquence de décalage du cadre de lecture au niveau des codons ANN et un (EL YACOUBI ET AL., 2011, DAUGERON ET AL., 2012). Cela suggère que seuls certains gènes doivent être particulièrement affectés par l'absence de t<sup>6</sup>A au niveau des ARNt. Le lien entre les phénotypes observés chez les mutants KEOPS/DEZ/Qri7/Sua5/YrdC et les conséquences de l'absence de t<sup>6</sup>A requiert donc une recherche systématique de ces gènes en question parmi les milliers de gènes présents dans le génome des Eucaryotes, des Bactéries et des Archées. Une approche a été initiée *in silico* (basé sur la richesse en codon ANN) (EL YACOUBI ET AL., 2011) et pointe une multitude de candidats potentiels.

Or, le génome mitochondrial de *S. cerevisiae* ne contient que 30 gènes qui sont traduits par la machinerie de traduction mitochondriale (FOURY ET AL., 1998). Cet organe représente donc un modèle simplifié pour étudier les conséquences traductionnelles liées à l'absence de t<sup>6</sup>A. Notre étude démontrant l'implication directe de Qri7 dans la biosynthèse de ce nucléoside, il sera maintenant possible de faire un lien entre les phénotypes observés par la

délétion de cette protéine et les gènes mitochondriaux dont l'expression est affectée par l'absence de t<sup>6</sup>A. Cette compréhension pourra servir de modèle pour cibler de manière plus précise les gènes dont l'expression dépend de la présence de t<sup>6</sup>A chez les Eucaryotes et chez les Bactéries, respectivement.

Chez l'Homme, de nombreuses maladies graves, notamment neurodégénératives, sont associées à des dysfonctionnements de la mitochondrie (appelées maladies mitochondriales), mais leur origine exacte n'est pas toujours connue (HOWEL, ET AL., 1999). Un très grand nombre de ces maladies sont associées à des mutations ponctuelles au niveau des ARNt mitochondriaux et récemment, il a été montré que deux pathologies mitochondriales majeures étaient causées par une telle mutation, conduisant à l'absence de la modification 5-taurinométhyluridine en position 34 de l'ARNt<sup>Leu</sup>, provoquant des erreurs de traduction au niveau de la mitochondrie (SUZUKI ET AL., 2011). Connaissant le rôle essentiel de t<sup>6</sup>A dans la traduction, il est très probable que certaines mutations au niveau des ARNt mt conduisant à l'absence de t<sup>6</sup>A puisse mener à des pathologies similaires. Il est donc indispensable de comprendre les déterminants moléculaires au niveau de l'ARNt qui permettent à Qri7 de reconnaître son substrat. Cette compréhension pourrait permettre d'identifier les mécanismes moléculaires à l'origine de certaines pathologies mitochondriales. La recherche de mutations au niveau d'*OSGL* (*Qri7*), qui pourraient conduire à une protéine non fonctionnelle ou anormale et entraîner un défaut de t<sup>6</sup>A au niveau des ARNt mt, est également une voie à explorer dans la compréhension des maladies mitochondriales.

### Qri7 représente un modèle simplifié pour l'étude de la machinerie de la synthèse de t<sup>6</sup>A.

Nos résultats indiquent que la protéine Qri7 réalise à elle seule la seconde étape de la réaction, à savoir le transfert du TC-AMP sur l'ARNt. Or, les machineries Bactérienne, Eucaryote et Archéenne qui réalisent cette étape (respectivement DEZ et KEOPS) sont des assemblages multiprotéiques, dépendant de l'interaction mutuelle de leurs composants, et qui présentent une activité d'hydrolyse de l'ATP, rendant la résolution du mécanisme catalytique complexe (DEUTSCH ET AL., 2012, LAUHON ET AL., 2012, PERROCHIA ET AL., 2013, ET PARTIE II)

La protéine Qri7 peut donc servir de modèle simplifié pour la compréhension de certaines modalités de ce mécanisme catalytique, grâce à sa caractérisation structurale et biochimique. Nous avons ainsi montré que Qri7 n'hydrolyse pas l'ATP, confirmant que cette activité n'intervient pas directement dans le mécanisme catalytique de formation de t<sup>6</sup>A lors de la seconde étape de la réaction. Ainsi, comme nous l'avons suggéré précédemment (PARTIE II), l'hydrolyse de l'ATP par KEOPS/DEZ doit jouer un rôle dans la régulation de leur activité (formation des interactions interprotéiques pour DEZ, interaction avec l'ARNt pour KEOPS).

Le transfert du TC-AMP implique également que Qri7 puisse à elle seule fixer l'ARNt. De manière surprenante, nos premiers essais en retard sur gel indiquent que Qri7 présente une faible affinité pour les ARNt. Cela suggère que Qri7 ne forme pas nécessairement un complexe ARNt/protéine stable et isolable par retard sur gel dans ces conditions, mais forme plutôt une association dynamique et transitoire. En revanche, dans les mêmes conditions, le complexe KEOPS de *P. abyssi* se fixe très fortement aux ARNt et forme un complexe nucléo-protéique stable et isolable (Perrochia et al, 2013). Nous avons suggéré précédemment (PARTIE II) que l'hydrolyse de l'ATP par Bud32, le partenaire de Kae1, ait pour rôle le relargage de l'ARNt. Le fait que Qri7 n'ait pas de partenaire équivalent et forme une association transitoire avec les ARNt semble confirmer notre hypothèse. Au contraire, YgjD présente une affinité comparable à celle de Qri7 en retard sur gel. Il se pourrait donc que la machinerie bactérienne interagisse également de manière transitoire avec l'ARNt, contrairement aux machineries Eucaryote/Archéenne. Des mécanismes différents vis-à-vis de l'interaction avec l'ARNt semblent exister entre les différentes machineries de synthèse de t<sup>6</sup>A, et l'étude comparative des machineries pourrait permettre de les comprendre.

La question de la discrimination des ARNt est également un point crucial. Tout comme pour la fixation, la protéine Qri7 doit pouvoir à elle seule reconnaître les ARNt substrats. Nous avons montré que Qri7 peut modifier au moins deux espèces d'ARNt différents. Il est nécessaire de tester d'autres types d'ARNt, notamment mitochondriaux, pour évaluer l'efficacité de la réaction en fonction du substrat et déterminer précisément ses modalités de reconnaissance. Il serait alors également intéressant d'identifier, par étude structurale, mutationnelle ou cross-linking, les résidus engagés dans l'interaction entre Qri7 et l'ARNt, ce



qui permettrait d'établir un modèle extrapolable pour les orthologues YgjD et Kae1 s'il s'agit de résidus conservés.

Cependant, on peut également suggérer que les déterminants reconnus par Qri7 ne sont pas les mêmes que ceux reconnus par les machineries bactériennes ou eucaryote et archéennes. En effet, l'ARNt<sup>Ser(GCU)</sup> mitochondrial de mammifère (dont l'Homme) ne présente pas de boucle D, alors qu'il porte t<sup>6</sup>A (ARCARI ET AL., 1980). Or, chez les Eucaryotes (le Xénope), il a été montré que la présence de cette boucle est nécessaire à la reconnaissance de la machinerie *cytoplasmique* (MORIN ET AL., 1998). La protéine Qri7 pourrait donc avoir au moins en partie des modalités de reconnaissance différentes des autres machineries.

### Qri7, une version de YgjD améliorée

L'étude phylogénétique montre que Qri7 présente une assez grande proximité avec YgjD, en raison de son origine évolutive (HECKER ET AL., 2007). Nous avons montré que la protéine Qri7 présente également au niveau biochimique des points communs avec YgjD par rapport à Kae1 : elle se trouve sous forme dimérique en solution comme YgjD (HANDFORD ET AL., 2009, ET DONNEES NON PUBLIEES), ne semble pas présenter de fer au sein de son site actif (YgjD comporte un atome de zinc qui n'est pas détectable par spectrométrie), et présente une affinité en gel retard comparable. On peut donc se demander quelles sont les caractéristiques qui permettent à Qri7 de fonctionner seule, alors que YgjD nécessite la présence stricte de ses partenaires. Pour répondre à cette question, il serait intéressant de comparer les séquences des différents orthologues de Qri7 de manière exhaustive chez les Eucaryotes (dont la liste a été dressée par OBERTO ET AL., 2009) afin de déterminer particularités de séquence qui les différencient du groupe d'orthologues des protéines YgjD. Si de telles différences sont identifiables, il serait alors envisageable par différentes mutations de « transformer » une protéine YgjD en *pseudo-Qri7* pouvant fonctionner sans l'aide de ses partenaires (Z et E). De manière intéressante, certains Mycoplasmes semblent ne pas posséder d'orthologue des protéines YjeE et YeaZ (OBERTO ET AL., 2013). On peut alors se demander si chez ces organismes, la protéine YgjD fonctionne seule, à l'instar de Qri7 chez la mitochondrie. Ainsi, l'étude de ces systèmes minimalistes permettrait de mieux comprendre le rôle des partenaires de YgjD, qui reste encore incompris, et de reconstituer le scénario évolutif des machineries de synthèse de t<sup>6</sup>A, dont nous parlerons par la suite.

## La fin des débats sur la complémentation entre YgjD et Qri7 chez *E. coli* ?

Avant que ne soit déterminé le rôle direct de YgjD et Qri7 dans la biosynthèse de t<sup>6</sup>A, des expériences de complémentation génétique ont été menées chez *E. coli* pour savoir si ces deux orthologues étaient fonctionnellement équivalents. La protéine YgjD étant essentielle chez *E. coli* (ARIGONI ET AL., 1998), une méthode d'inactivation conditionnelle de ce gène est nécessaire. Oberto et al et El Yacoubi et al ont utilisé respectivement deux systèmes génétiques différents pour mener ces expériences, en inactivant conditionnellement le gène *ygjD* et en apportant une copie du gène *qri7*. Dans le système de Oberto et al (système A), l'expression de *qri7* complémente l'inactivation de *ygjD* (OBERTO ET AL., 2009) ; au contraire, El yacoubi et al n'observent pas complémentation de YgjD par Qri7 dans leur système (système B). (EL YACOUBI ET AL., 2011). Ces systèmes présentent deux différences majeures, qui peuvent chacune fournir des explications à ces résultats opposés:

- dans le système A, le gène *ygjD* est sous contrôle du promoteur pBAD, alors que dans le système B, *ygjD* est sous contrôle du promoteur pTET. Il a été suggéré que la complémentation observée dans le système A soit due à une fuite du promoteur pBAD, moins stringente que l'inactivation induite par le promoteur pTET (EL YACOUBI ET AL., 2011)
- dans le système A, la protéine Qri7 exprimée est délétée de ses 27 premiers acides aminés (correspondant au signal d'adressage mitochondrial), alors que dans le système B, la version de Qri7 est délétée de 30 acides aminés. Il se pourrait donc que la présence de 3 acides aminés supplémentaires dans le système B engendre une protéine non fonctionnelle, expliquant l'absence de complémentation observée.

La protéine Qri7 que nous avons purifiée correspond à la version employée par Oberto et al, délétée de 30 premiers acides aminés. Nous avons montré que cette protéine, *in vitro*, forme un système de biosynthèse de t<sup>6</sup>A viable avec YrdC à 37°C, dont l'efficacité est comparable à la machinerie bactérienne complète. Ce résultat suggère donc que Qri7 exprimée dans *E. coli* est potentiellement capable de modifier les ARNt en t<sup>6</sup>A, donc de remplacer fonctionnellement la perte de YgjD, et confirme le résultat d'Oberto et al. Pour expliquer le résultat négatif du système génétique de El yacoubi et al., il serait nécessaire de purifier une protéine délétée de ses 27 premiers acides aminés et de vérifier sa capacité à catalyser la synthèse de t<sup>6</sup>A.

Ces expériences de génétique avaient pour but initial de comprendre le rôle des protéines YgjD/Qri7 et leur équivalence. Grâce aux différentes études biochimiques (DEUTSCH ET AL., 2012, LAUHON ET AL., 2012 ET NOS RESULTATS), il a été prouvé YgjD et Qri7 ont effectivement le même rôle (catalyse de la seconde étape de biosynthèse de t<sup>6</sup>A), mais ne fonctionnent pas de la même manière. Le débat engendré par les expériences de complémentation relève donc probablement de différences techniques et n'apparaît plus aussi déterminant dans la compréhension de la fonction de Qri7 et YgjD.

### Une machinerie de synthèse de t<sup>6</sup>A chez les plastes ?

Les plastes sont une autre catégorie d'organite présents chez différents groupes d'Eucaryotes (algues brunes, algues rouges, lignée verte, autres eucaryotes unicellulaire). Les plastes de la lignée verte, tout comme la mitochondrie, sont issus d'une endosymbiose primaire entre un Eucaryote (possédant déjà une mitochondrie) et une bactérie, semblable aux cyanobactéries actuelles (SAGAN L. 1967). Du fait de cette origine et de ce statut d'organite, les plastes présentent de nombreux points communs avec les mitochondries. Ils ont été soumis à une évolution réductrice comparable, caractérisé par une perte de gènes, un transfert de gène vers le génome nucléaire, et la mise en place de mécanisme de réadressage. Le plaste de l'espèce modèle *Arabidopsis thaliana* possède un génome très réduit : 154 kb, codant 87 gènes protéiques, 4 ARNr, et un set complet de 30 ARNt représentant les 20 acides aminés du code génétique (SUSHEI ET AL., 1999). Ces organites possèdent une machinerie de traduction fonctionnelle de type bactérien (LYSKA ET AL. 2012), et les banques de données indiquent que la modification t<sup>6</sup>A est présente au niveau des ARNt plastidiens (MACHNICKA ET AL., 2013 -MODOMICS). On peut donc s'attendre à la présence d'une machinerie de synthèse de t<sup>6</sup>A au sein des plastes. Certaines données protéomiques semblent indiquer la présence d'orthologue de YrdC et de YgjD/Kae1 au niveau des plastes (PLANT PROTEOMICS DATABASE , SUN ET AL., 2009), mais aucune étude établissant clairement cette présence n'a été publiée pour le moment. Ainsi, la machinerie t<sup>6</sup>A du plaste représente encore un autre cas de machinerie de synthèse de t<sup>6</sup>A qui reste à caractériser.

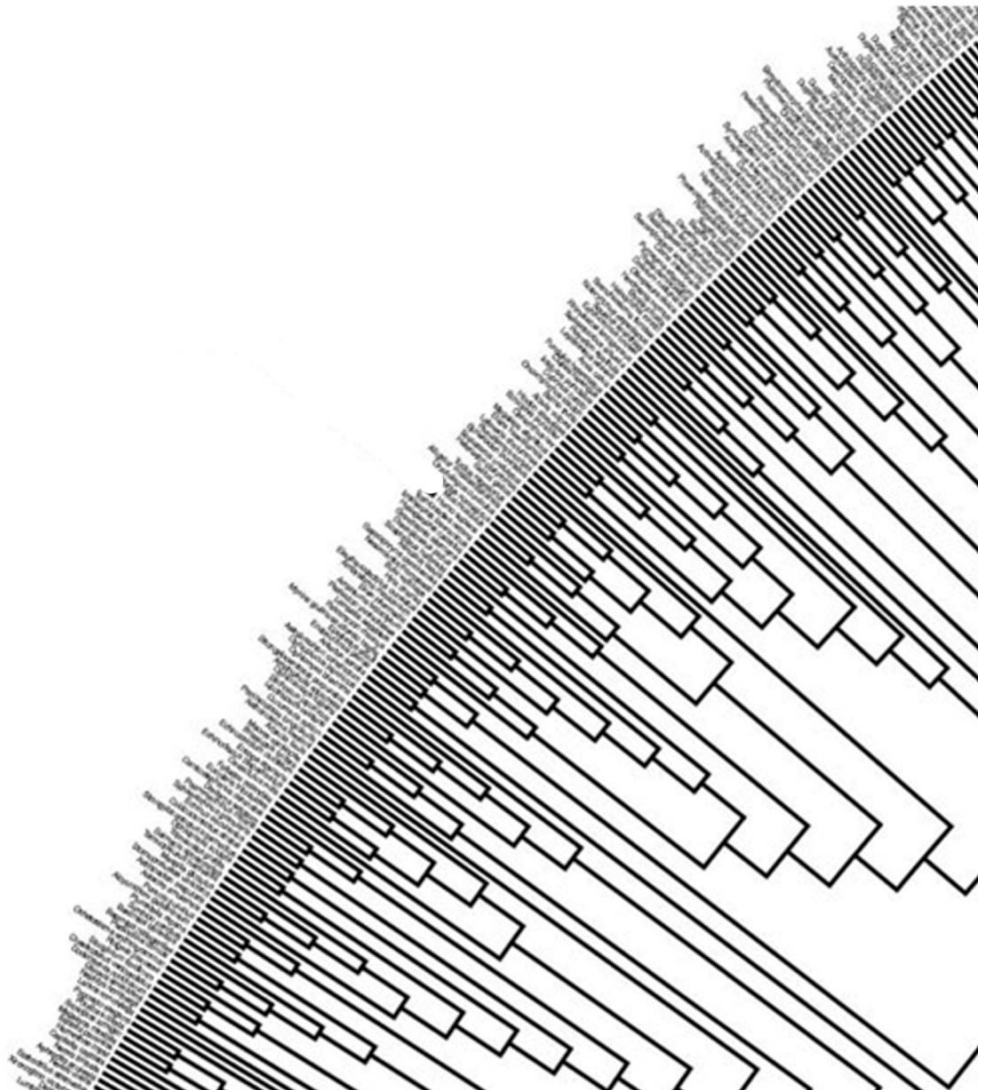
## Conclusion

La mitochondrie est au cœur de très nombreuses études, car elle joue un rôle majeur dans la cellule eucaryote, au niveau du métabolisme oxydatif, de diverses voies métaboliques (cycle de l'urée par exemple) mais aussi les mécanismes d'apoptose. De très nombreuses maladies, notamment des maladies neurodégénératives graves sont associées à des dysfonctionnements de cet organite, et plus précisément à des mutations au niveau des ARNt mitochondriaux. Nos résultats ouvrent des voies potentielles quant à la compréhension de l'origine moléculaire de certaines de ces pathologies, qui pourraient être liée à l'absence  $t^6A$ . La caractérisation détaillée des modalités de fonctionnement de sa machinerie de synthèse est donc un enjeu primordial.

La mitochondrie, en tant que descendante d'une bactérie, est également très étudiée d'un point de vue évolutif. De manière intéressante, l'évolution réductrice qu'elle a connue semble se refléter dans sa machinerie de biosynthèse de  $t^6A$ , qui ne comprend aujourd'hui que deux composants, les protéines Qri7 et Sua5, alors que les autres machineries fonctionnent avec au minimum 4 protéines différentes. Le maintien d'une machinerie de synthèse de  $t^6A$ , même réduite, traduit l'importance majeure que doit jouer ce nucléoside dans tous les systèmes de traduction actuels. Cependant, cette évolution réductrice ne signifie pas que la protéine Qri7 est « dégénérée » par rapport à son orthologue bactérien : au contraire, nos résultats montrent qu'elle est capable de catalyser à elle seule une réaction qui nécessite la présence stricte de plusieurs partenaires dans les autres systèmes. Ainsi, l'étude de cette machinerie particulière permet non seulement de servir de modèle simplifié du point de vue biochimique, mais permet aussi de proposer et d'imaginer des scénarios intéressants quant à l'histoire évolutive des différents systèmes de synthèse de  $t^6A$ .

Nous allons maintenant aborder ces aspects évolutifs dans une conclusion générale, après bref rappel de l'ensemble des résultats obtenus au cours de cette thèse, pour finir sur les différentes questions et piste qui restent encore ouverte.

# Conclusion et Perspectives

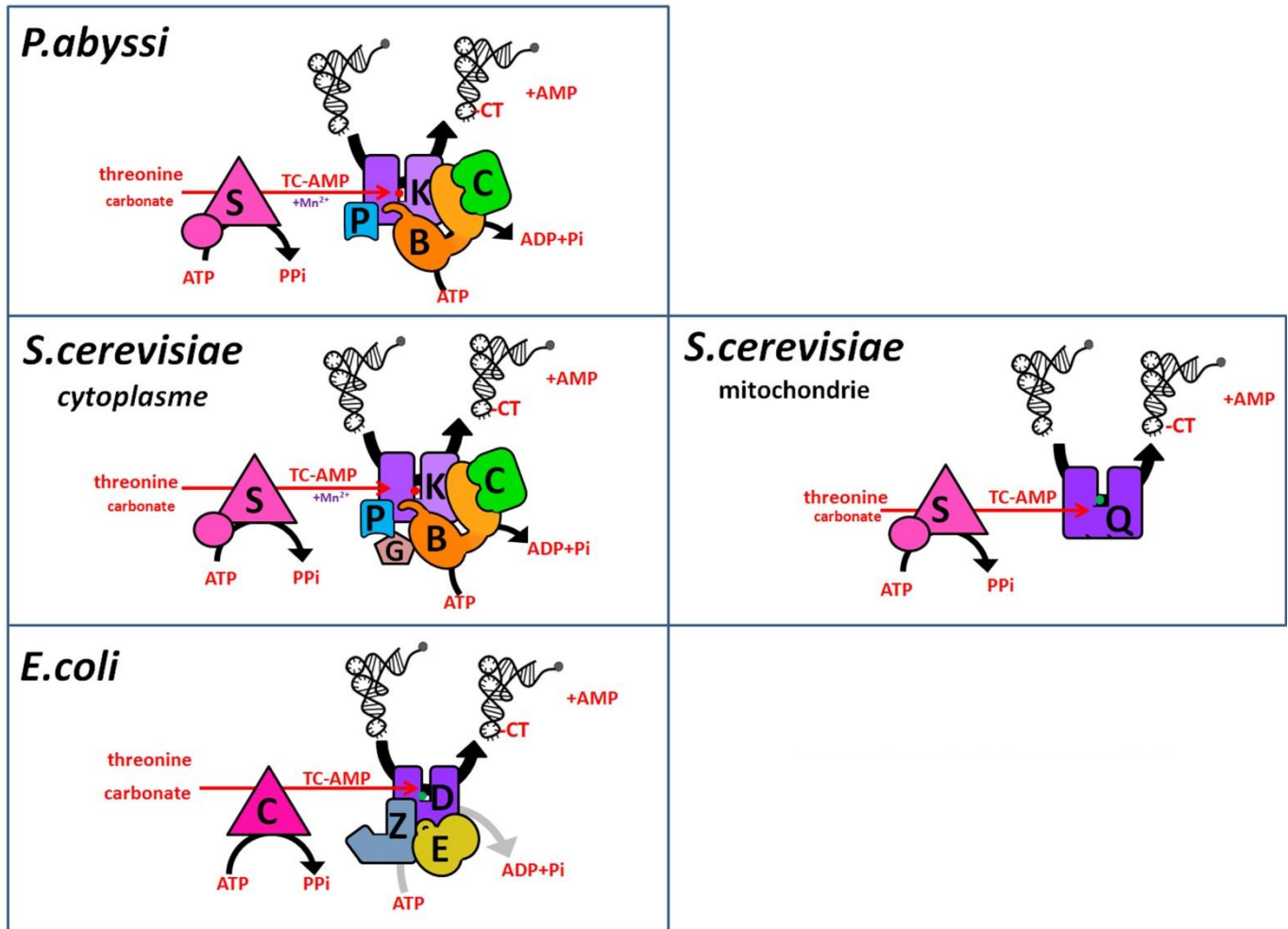




## Rappel de l'ensemble des résultats

Nous avons l'objectif de répondre à plusieurs questions. La première d'entre elles était la suivante: **chez les Archées et les Eucaryotes, quels sont les composants de la machinerie de synthèse de t<sup>6</sup>A ?** Les études génétiques avaient pointé le rôle de Sua5 et Kae1 chez *S.cerevisiae*, et le rôle probable des partenaires de Kae1 au sein du complexe KEOPS, sans pour autant apporter une démonstration directe. Nous avons donc choisi une approche biochimique *in vitro*, et avons montré que cette machinerie est formée de la protéine Sua5 et du complexe KEOPS chez *P. abyssi* et *S. cerevisiae*. Ce travail a permis d'éclairer le mécanisme de cette réaction, et de proposer un premier modèle de biosynthèse de t<sup>6</sup>A chez ces organismes. Nous avons pu montrer que les deux composants (Sua/YrdC d'une part, et KEOPS/DEZ d'autre part) issu des trois machineries Archéenne, Bactérienne, et Eucaryote sont interchangeable *in vitro*, indiquant d'une part que les orthologues de Sua5/YrdC ont le même rôle, et que le complexe KEOPS et le complexe DEZ sont des analogues fonctionnels.

Ayant montré que Kae1 ne suffit pas à elle seule à synthétiser t<sup>6</sup>A aux côtés de Sua5, nous avons donc tenté de répondre à la question suivante : **quel est le rôle des sous-unités du complexe KEOPS ?** Nous avons montré que chacune des sous-unités de KEOPS, chez *P. abyssi*, est nécessaire pour une activité optimale, et qu'un réseau d'interactions mutuelles est à l'œuvre au sein de ce complexe. Basé sur nos résultats et des données disponibles dans la littérature, il apparaît que la protéine Kae1 est le site de fixation de l'intermédiaire TC-AMP produit par Sua5 et est responsable du transfert du groupement TC sur l'ARNt. Sa partenaire, la protéine Pcc1, est impliquée aux coté de Kae1 dans la fixation, voire dans la reconnaissance de l'ARNt. De son côté, la protéine Cgi121, facteur accessoire, semble jouer le rôle de chaperonne pour stabiliser l'ensemble du complexe, et activer sa partenaire : Bud32. Cette dernière protéine, régulée également par son interaction avec Kae1, est une ATPase dont l'activité d'hydrolyse est essentielle à l'activité de synthèse de t<sup>6</sup>A par le complexe KEOPS. Nos données suggèrent que Bud32 fonctionne comme un interrupteur moléculaire dont l'activité ATPase qui joue un rôle dans l'interaction entre KEOPS et le substrat ARNt. Ainsi, un mécanisme complexe et hautement contrôlé est à l'œuvre au sein de cette machinerie protéique pour effectuer la seconde étape de la réaction.



**FIGURE C-1** : Représentation schématique des différentes machineries de synthèse de t6A des espèces modèles *P. abyssi* (Archée), *S. cerevisiae* (Eucaryote) et *E. coli* (Bactérie)

S=Sua5, C(rose)= YrdC, K=Kae1, P= Pcc1, B= Bud32, C(vert)=Cgi121, Z=YeaZ, E=YjeE, Q= Qri7, G= Gon7 (uniquement chez les Champignons). Le rond rose sur S symbolise la présence du domaine SUA5-C terminal.

TC-AMP= thréonyl-carbamoyl-AMP, CT= N<sup>6</sup>thréonylcarbamoyl greffé sur l'adénosine 37 de l'ARNt

Le rond rouge au niveau de K représente l'atome de fer au sein du site actif, le rond vert au niveau de D (et Q) l'atome de zinc. L'effet stimulant du manganèse sur la réaction retrouvé pour les machineries de *P. abyssi* et *S. cerevisiae* (cytoplasmique) est indiqué. La flèche grise correspondant à l'hydrolyse de l'ATP dans le système *E. coli* indique que la protéine responsable de cette activité n'est pas déterminée.



Après avoir caractérisé cette machinerie présente chez les Archées et les Eucaryotes au niveau cytoplasmique, une question restait en suspens : **comment est synthétisée t<sup>6</sup>A chez la mitochondrie ?** Nos résultats, en accord avec des résultats de localisation (publiés pour Qri7 et non publiés pour Sua5) ont montré que chez cet organite, la machinerie de synthèse est formée de seulement deux protéines : la protéine Sua5 et la protéine Qri7, orthologue de YgjD/Kae1. Nous avons ainsi montré que cette machinerie mitochondriale est un système minimaliste, qui ne requiert pas l'intervention de partenaires spécifiques, et dont l'étude offre un modèle simplifié pour la compréhension du mécanisme de biosynthèse de t<sup>6</sup>A.

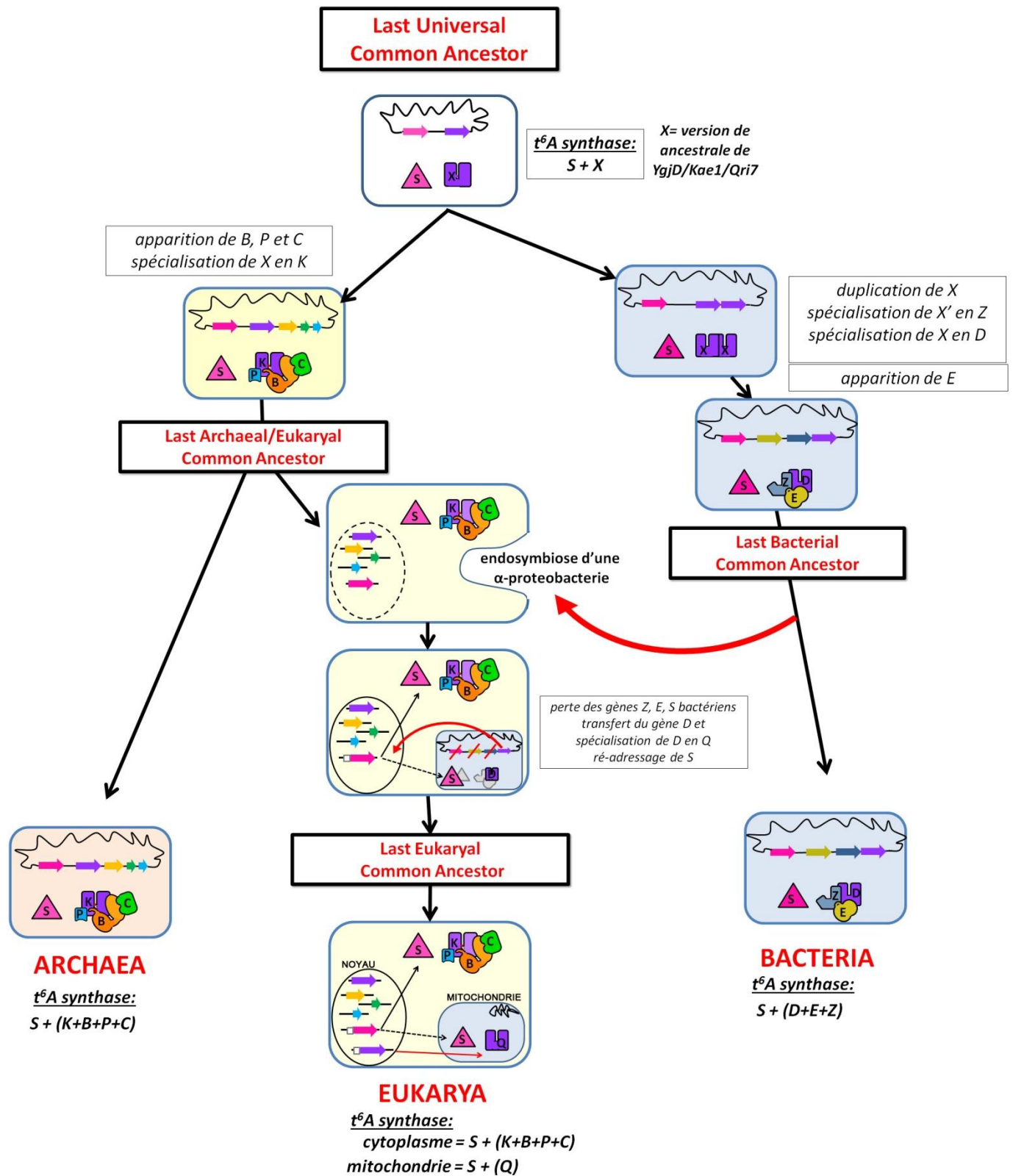
A l'issue de ces travaux, il est donc possible d'établir un parallèle entre les machineries de biosynthèse de t<sup>6</sup>A de l'Archée *Pyrococcus abyssi*, de l'Eucaryote *Saccharomyces cerevisiae*, des Bactéries *Escherichia coli* et *Bacillus subtilis*, et de la mitochondrie de *S.cerevisiae* (**FIGURE C-1**). Ce parallèle permet d'identifier les caractères universels qui unissent ces différentes machineries, mais aussi les éléments qui les distinguent, permettant de proposer des scénarios évolutifs pour retracer leurs histoires respectives et communes (**FIGURE C-2**)

## Scénario de l'histoire évolutive de la machinerie de synthèse de t<sup>6</sup>A

### Un système universel, mais une évolution convergente :

Depuis la découverte de t<sup>6</sup>A à la fin des années 1960, de nombreuses études structurales et biochimiques ont montré que la présence de ce nucléoside au sein de l'ARNt doit jouer un rôle majeur à la fois dans l'efficacité et la fidélité de la traduction, et l'identification récente de sa voie de biosynthèse permet maintenant de confirmer ce rôle *in vivo*. t<sup>6</sup>A<sub>37</sub> est l'une des rares modifications d'ARNt véritablement universelles, ce qui renforce l'idée qu'il s'agit d'un élément fondamental du système de traduction moderne.

La caractérisation des machineries de biosynthèse de t<sup>6</sup>A chez les différents domaines du vivant montre que ce sont deux protéines universelles, YrdC et Kae1/YgjD, qui sont responsables de cette réaction, mais que des partenaires spécifiques non universels sont strictement requis. Or, l'existence d'un système minimaliste fonctionnel à deux



**FIGURE C-2 :** Scénario hypothétique de l'évolution de la machinerie de biosynthèse de t<sub>6</sub>A dans les trois domaines du vivant

L'explication est donnée dans le texte. X= version ancestrale de K/D/Z/Q, S=Sua5/YrdC, K=Kae1, P= Pcc1, B= Bud32, C=Cgi121, Z=YeaZ, E=YjeE, Q= Qri7

composants [Sua5+Qri7] chez la mitochondrie indique qu'il est possible que le dernier ancêtre commun de tous les organismes, **LUCA, possédait lui aussi un système ancestral à deux composants uniquement [YrdC+Kae1/YgjD] qui était déjà *fonctionnel*. Basé sur ces données, nous pouvons donc proposer que LUCA avait déjà des ARNt portant le nucléoside t<sup>6</sup>A et, selon cette hypothèse, un système traductionnel semblable au système moderne, effectuant une traduction fidèle et efficace.**

L'apparition des autres composants aurait par la suite permis une amélioration de l'efficacité de la synthèse de t<sup>6</sup>A, renforçant la vitesse de catalyse, le rendement, et/ou la sélectivité des ARNt à modifier, et conférant un avantage sélectif certain. Ce scénario reflète ainsi un cas d'évolution convergente concernant les machineries de synthèse de t<sup>6</sup>A moderne: la protéine Kae1/YgjD aurait acquis au cours de l'évolution des partenaires non homologues dans les lignées Bactéries d'une part, et Eucaryotes/Archées d'autre part, permettant dans les deux cas une amélioration du système initial à l'aide de deux stratégies différentes.

L'étude comparative de ces machineries, et la compréhension exacte du rôle de chaque composant permet donc l'élaboration de scénarios sur l'histoire évolutive de cette voie de biosynthèse et de retracer l'état et le devenir des différentes stades ancestraux de ces machineries. Le scénario global que nous proposons peut être détaillé grâce à certaines observations et données dont nous disposons.

### « La simplicité est la sophistication extrême » (Leonard de Vinci) – le cas de la mitochondrie :

Toutes les Bactéries actuelles (à l'exception des Mollicutes, qui constituent un cas particulier) possèdent les 4 composants [YrdC, YgjD, YjeE et YeaZ] (OBERTO ET AL, 2013). On peut donc supposer que ces 4 protéines aient été présentes chez le dernier ancêtre commun bactérien (LBCA), qui a donné naissance aux différentes lignées modernes, dont les  $\alpha$ -protéobactéries. Or, les études de phylogénie moléculaires ont montrées que la mitochondrie des Eucaryotes actuels descend d'un même ancêtre issu de cette lignée bactérienne (Andersson *et al*, 2003). Ainsi, au moment du processus d'endosymbiose, l'ancêtre mitochondriale devait posséder également une machinerie à 4 composants.

Pourtant, toutes les mitochondries ne présentent actuellement plus que YgjD (évolué en Qri7) comme relique de cette machinerie (OBERTO ET AL., 2009)

La version YrdC a été probablement remplacée par son orthologue eucaryote, tandis que les protéines YeaZ et YjeE ont été perdues. Cette perte a dû se produire parallèlement à des modifications dans la séquence d'YgjD, lui permettant de fonctionner sans l'aide de partenaires. Ainsi, la mitochondrie traduit un phénomène de simplification, par un passage d'un système complexe (chez l'ancêtre) vers un système simplifié moderne et fonctionnel. Ce phénomène semble avoir également eu lieu chez certaines Mollicutes, où seules YrdC et YgjD sont détectables (OBERTO ET AL., 2013). La résistance de la machinerie de synthèse de t<sup>6</sup>A à la perte implique que ce nucléoside modifié, une fois acquis au cours de l'évolution, a dû représenter un avantage sélectif majeur et devenir strictement indispensable au fonctionnement correcte de la machinerie de traduction de la cellule.

Le fait que toutes les mitochondries ne possèdent plus que YgjD (transformée en Qri7) suppose également que les événements de perte de YeaZ et YjeE et d'adaptation de YgjD se soient produits avant la divergence des lignées Eucaryotes actuelles, confortant l'idée qu'un temps relativement long s'est déroulé entre l'évènement d'endosymbiose primaire et la divergence des Eucaryotes, comme l'étude du ribosome mitochondrial le suggère. (DESMOND ET AL., 2011.).

### Chez les Bactéries : d'un homodimère à un hétérodimère

La protéine YgjD et YeaZ actuelles forment chacune des homodimères en solution (HANDFORD, ET AL., 2009, ET DONNEES NON PUBLIEES), et la protéine YeaZ est un paralogue d'YgjD (HECKER ET AL. 2007), qui a perdu certains éléments structuraux de son site catalytique (NICHOLS ET AL., 2005, HECKER ET AL, 2007). De plus, l'hétérodimère YeaZ/YgjD présente une interface spécifique d'une espèce à l'autre (RAJAPOLA ET AL., 2011) et les expériences de complémentation montrent que ces deux protéines se comportent comme une seule unité fonctionnelle (EL YACOUBI ET AL., 2011). Nous pouvons donc proposer le scénario suivant : chez la lignée ayant donné naissance à LBCA (l'ancêtre commun à toutes les lignées bactériennes actuelles), la protéine YgjD fonctionnait seule sous forme d'homodimère (aux côtés de YrdC

pour la première étape). Le gène codant cette protéine aurait alors été dupliqué, et l'une des versions aurait divergé, n'étant plus soumis à la pression de sélection, pour aboutir à la protéine YeaZ dépourvu d'activité catalytique. Cette nouvelle version aurait conservé la capacité de se dimériser, permettant la formation d'un couple hétérodimérique YeaZ/YgjD au sein de la cellule. Ce couple aurait pu être plus efficace qu'un couple homodimérique pour la synthèse de t<sup>6</sup>A, et aurait alors co-évolué en une unité fonctionnelle dans chaque lignée bactérienne, renforçant la spécificité de leur interface. La troisième protéine, YjeE, qui a du également apparaître avant la divergence des lignées bactériennes, a pu être recrutée par le couple DZ et fournir un mécanisme de contrôle ou de régulation, et augmentant encore l'efficacité globale du système de synthèse de t<sup>6</sup>A.

### Du côté des Archées et des Eucaryotes...

Presque tous les génomes Eucaryotes et Archées présentent les gènes identifiables du complexe KEOPS actuel Pcc1, Bud32, Cgi121. De plus, ces composants interagissent de la même façon chez *S. cerevisiae*, *H. sapiens* et *P. abyssi*, des organismes phylogénétiquement éloignés. Cela implique que l'apparition de ces partenaires et la spécialisation de leurs interactions sont des événements très anciens, survenus avant la divergence des lignées ayant donné naissance d'une part aux Archées et d'autre part aux Eucaryotes. Le recrutement de ces partenaires par Kae1 aurait également, à l'image du système Bactérien, permis d'améliorer l'efficacité de synthèse de t<sup>6</sup>A.

De manière intéressante la protéine Kae1 de *P. abyssi* ne forme pas de dimère (HECKER ET AL. 2007), contrairement aux sous-complexe Pcc1-Kae1 et au complexe KEOPS complet fonctionnel (NOS DONNEES, MAO ET AL, 2008, COSTESSI ET AL., 2012). Chez les autres machineries actuelles, on constate que Qri7, qui fonctionne seule, est dimérique en solution, et le couple YgjD/YeaZ est un hétérodimère, probablement comme on l'a vu issu d'un état homodimérique ancestral. Il semble donc que l'état dimérique soit une caractéristique commune à ces machineries: la protéine YgjD/Kae1 de LUCA était elle également dimérique ? Si c'est le cas, chez les Archées et les Eucaryotes, la protéine Kae1 aurait alors perdu sa capacité à se dimériser seule, Pcc1 jouant ce rôle. Cette protéine aurait donc pu

être recrutée par sa capacité à se dimériser, mais aussi peut être pour son affinité pour les ARNt.

L'histoire évolutive de Bud32, enfin, est également intéressante. Cette protéine, classée comme une kinase atypique (STOCCHETTO ET AL, 1997), est très proche d'une autre famille de kinase atypique, les Rio kinases (Leonard et al, 1998). Cette famille est retrouvée chez les Eucaryotes, chez les Archées mais aussi chez les Bactéries (LARONDE ET AL, 2005, ESSER ET AL, 2013) Plusieurs caractéristiques unissent Bud32 et les Rio: elles sont dépourvues de certaines éléments caractéristiques des kinases eucaryotes actuelles (qui permettent notamment la reconnaissance du substrat), mais peuvent néanmoins s'autophosphoryler *in vitro* et phosphoryler de manière aspécifique certains substrat. De plus, elles sont impliquées dans le métabolisme des ARN : ARNt pour le cas de Bud32, particule ribosomale dans le cas des Rio (ESSER ET AL, 2013). Récemment, il a été montré que Rio2 fonctionne en réalité comme une ATPase (FEIRREIRA CERCA, 2012), dont l'hydrolyse de l'ATP assure un contrôle de la maturation. Nos résultats indiquent un comportement similaire pour Bud32, et l'activité phosphotransférase observée *in vitro* ne semble donc pas liée à la fonction biologique de ces protéines, et pourrait être seulement une activité « aspécifique ».

Ces différents éléments permettent de proposer un scénario évolutif pour cette grande famille de protéine. Chez LUCA, une version ancestrale d'ATPase-Rio existait déjà, impliquée dans la maturation du ribosome. Chez les Bactéries, cette version a donné naissance aux Rio actuelles bactériennes. Chez la lignée Archées/Eucaryotes, cette ATPase ancestrale aurait été dupliquée pour donner naissance à deux familles : d'une part Bud32, qui aurait alors été recrutée spécifiquement par Kae1 au niveau des ARNt dans la voie de synthèse de t<sup>6</sup>A, apportant un élément de régulation avantageux, et d'autre part les Rio, qui auraient conservé leur rôle dans la biogénèse du ribosome. Enfin, dans la lignée Eucaryote spécifiquement, une sous-famille de Rio kinase aurait acquis des nouveaux éléments structuraux, permettant de rendre l'activité phosphotransférase spécifique et fonctionnelle, et donner naissance aux kinases eucaryotes actuelles.

Ces scénarios évolutifs proposés sont basés sur l'observation des systèmes actuels. Ne disposant pas de machines à remonter dans le temps, ils ne représentent qu'une des histoires possibles et envisageables. Nous avons proposé une machinerie minimale de

biosynthèse de t<sup>6</sup>A chez LUCA avec deux composants, YrdC et Kae1. Il est possible que ce système ait été en réalité déjà plus complexe, comportant d'autres composants. Par exemple, les membres du complexe KEOPS aurait pu être présents aux cotés de Kae1, et perdus par la suite dans la lignée ayant donnée naissance aux Bactéries pour être remplacé par les protéines YeaZ et YjeE. Afin de renforcer les scénarios les plus probables, la caractérisation détaillée des machineries, de leur fonctionnement, et du rôle exact de chaque composant doit donc être poursuivie.

## La biosynthèse de t<sup>6</sup>A, entre universalité et variabilité

A l'issue de ces travaux, il est donc possible d'établir un parallèle entre les machineries de biosynthèse de t<sup>6</sup>A de l'Archée *Pyrococcus abyssi*, de l'Eucaryote *Saccharomyces cerevisiae*, des Bactéries *Escherichia coli* et *Bacillus subtilis*, et de la mitochondrie de *S.cerevisiae*. Ce parallèle permet d'identifier les caractères universels qui unissent ces différentes machineries, mais aussi les éléments qui les distinguent, ouvrant la voie à de nouvelles questions et de très nombreuses pistes d'études (**FIGURE C-2**)

### Le premier composant Sua5/YrdC : une activité de production de TC-AMP universelle, mais une variabilité structurale intrigante

Le premier composant de chacune des machineries est l'orthologue de la famille Sua5/YrdC, qui produit un intermédiaire réactionnel TC-AMP, dont l'existence a été démontrée pour les systèmes bactériens et archéens (**LAUHON ET AL, 2012, PARTHIER ET AL. 2012**). Au vue des expériences de complémentation que nous avons effectuées, et de la similarité de séquence de tous les orthologues au niveau de leur domaine catalytique YrdC, il est très probable que cet intermédiaire soit produit de manière universel chez tous les êtres vivants. **Le décryptage exact du mécanisme de sa formation, à partir d'ATP, de carbonate libre, et de thréonine, reste encore à établir et permettra de savoir si les différents orthologues présentent un mécanisme identique.**

Malgré cette universalité fonctionnelle, trois « modèles » différents d'orthologue ont été mis en évidence : un modèle « YrdC », ne comprenant que le domaine catalytique (chez certaines Bactéries et Archées), un modèle « Sua5 », où le domaine YrdC est fusionné avec un domaine C-terminal Sua5 (chez certaines Bactéries, Archées et Eucaryotes) et un modèle « IRIP », où le domaine catalytique YrdC est inclus au sein d'un polypeptide de plus grande taille (chez l'Homme et les mammifères notamment). La présence de domaines supplémentaires au domaine catalytique pourrait moduler de manière directe l'activité de celui ci, comme semblent le suggérer nos données concernant la protéine de *Pyrococcus abyssi*, et donc pointer des différences de fonctionnement entre les différents orthologues. Il



est également possible que ces domaines soient des éléments régulateurs permettant de diriger YrdC vers une autre fonction que la synthèse de t<sup>6</sup>A. **Il reste donc à comprendre le rôle et l'activité de ces extra-domaines.**

De plus, la répartition de ces différents « modèles » d'orthologues ne semble pas correspondre à la phylogénie des organismes, ce qui suggère des phénomènes de multiples transferts latéraux de gènes entre Bactéries, Archées et Eucaryotes. L'étude phylogénétique de ces orthologues est donc fondamentale **pour comprendre l'histoire évolutive de cette famille universelle**, et remonter éventuellement à la forme ancestrale de la protéine YrdC.

#### **Intérêt pour la recherche appliquée :**

*La compréhension exacte du rôle et du fonctionnement du domaine additionnel C-terminal SUA5 présente également un intérêt en terme de santé publique, car ce domaine est absent chez l'Homme mais retrouvé fréquemment chez les parasites eucaryotes unicellulaires, notamment chez les Trypanosomes, les agents responsables de la leishmaniose et du paludisme (d'après la banque de données PFAM, Figure I-1, Partie I). On le retrouve systématiquement chez les champignons, dont certains sont également responsables de maladies chez l'Homme mais aussi chez de nombreuses plantes. La caractérisation de ce domaine protéique pourrait ainsi mener à terme à la découverte de nouvelles stratégies anti-parasitaires et anti-fongiques, basé sur le développement de molécules spécifiques inhibant son activité.*

#### **Le second composant Kae1/YgjD : une activité universelle de transfert du TC sur l'ARNt**

En plus de l'orthologue Sua5/YrdC, toutes les machineries biosynthèse de t<sup>6</sup>A décrites ont pour point commun la seconde protéine universelle du système, membre du COG0053 (Kae1/YgjD/Qri7). Les différentes études biochimiques et structurales indiquent que cette protéine est le lieu de fixation de l'intermédiaire TC-AMP produit par Sua5/YrdC.

D'un point de vue métabolique, l'instabilité de cet intermédiaire *in vitro* (Lauhon et al, 2012) suggère l'existence d'un couplage dans la cellule permettant de coordonner l'activité des deux composants de la machinerie. Un tel couplage existe dans les carbamoyl-transférases où les domaines YrdC et Kae1 sont fusionnés (Petkun et al, 2011, Parthier et al. 2011). Dans la machinerie bactérienne, un couplage similaire semble être envisageable, car une interaction directe a été décrite entre YrdC et YgjD (DEUTSCH ET AL. 2012), bien qu'il ne semble pas strictement indispensable au fonctionnement du système (LAUHON 2012). Dans les machineries Eucaryotes et Archées, aucune interaction de ce type n'a été détectée. **Ainsi, les modalités de transition de l'intermédiaire réactionnel de Sua5/YrdC vers Kae1/YgjD reste une question ouverte pour comprendre précisément le fonctionnement de ces machineries complexes.**

Une fois l'intermédiaire fixé à Kae1, le groupement TC doit être transféré sur l'ARNt. Cette étape nécessite la présence d'un ion métallique présent systématiquement au niveau du site catalytique de tous les orthologues, coordonnée par deux histidines conservées. Pourtant, les orthologues Kae1 (Eucaryotes et Archées) diffèrent des versions Bactériennes par la nature de cet ion divalent : un atome de Fer dans le premier cas (HECKER ET AL, 2007), un atome de zinc dans le second cas (NICHOLS ET AL, 2013), la nature de l'ion présent au niveau de Qri7 restant à définir.

De manière intéressante, la structure de la thréonyl-ARNt synthétase (une aaRS de classe II) cristallisée avec un analogue du thréonyl-AMP (l'intermédiaire réactionnel) révèle la présence d'un atome de zinc au sein de son site actif, coordonné par deux histidines. Cet atome n'a pas un rôle strictement structural ou catalytique, mais joue un rôle direct dans la spécificité pour la thréonine, en coordonnant son groupement OH, et permet d'éviter l'incorporation d'un autre acide aminé (SANKARANARAYANAN ET AL., 2000). Ce mécanisme de discrimination atypique n'est présent chez aucune autre aaRS. Par analogie, le métal du site actif de Kae1/YgjD/Qri7 pourrait donc jouer le même rôle dans le contrôle de la reconnaissance de l'AMP-TC, et traduirait ainsi un cas de convergence évolutive.

La compréhension du rôle précis de cet ion et de sa nature (Fer ou Zinc) pourrait ainsi révéler certaines spécificités au niveau mécanistique entre les orthologues Kae1/YgjD/Qri7, qui catalysent cette réaction universelle de transfert.

## KEOPS et DEZ : des analogues fonctionnels... mais des fonctionnements différents

Alors que le centre catalytique de toutes les machineries, Kae1/YgjD, est universellement conservé, le nombre et la nature des partenaires associés diffèrent d'un système à l'autre. Chez les Bactéries, deux partenaires spécifiques sont requis : YeaZ et YjeE (DEUTSCH ET AL, 2012, LAUHON ET AL, 2012), qui forment un complexe dynamique (appelé DEZ) (NICHOLS ET AL, 2013), alors que chez les Archées et les Eucaryotes, trois partenaires interviennent : Pcc1, Bud32 et Cgi121, qui forment le complexe KEOPS. Les Champignons, dont fait partie *S.cerevisiae*, sont un cas particulier, car un 5<sup>ème</sup> élément est présent : la protéine Gon7. Enfin chez la mitochondrie, aucun partenaire n'est présent et nécessaire. Cette dernière information indique que les partenaires spécifiques ne sont pas strictement essentiels au mécanisme catalytique, et sont des éléments de régulation, de contrôle de l'efficacité de la réaction ou de la discrimination des substrats.

Nous avons montré que le complexe KEOPS (Archées et Eucaryotes) et le complexe bactérien DEZ sont des analogues fonctionnels interchangeables. De plus, ils présentent chacun une activité d'hydrolyse de l'ATP, qui n'intervient pas directement dans le mécanisme catalytique de la réaction, étant donné que la protéine Qri7 ne fait pas appel à cette hydrolyse.

Chacun des deux systèmes possède une protéine à P-loop: Bud32 et YjeE, respectivement. Nous avons montré que Bud32 est bien responsable de l'hydrolyse de ADP +Pi qui est liée à l'interaction entre l'ARNt et KEOPS. De manière intéressante, la structure de YjeE suggère que cette protéine fonctionne comme un interrupteur moléculaire via son activité ATPase (TEPYAKOV ET AL., 2002). Ainsi, il semble que de manière indépendante, les

protéines Kae1 et YgiD ait fait appel à un mécanisme de *switch moléculaire* contrôlé par l'hydrolyse de l'ATP.

Cette activité d'hydrolyse de l'ATP ne joue pourtant pas un rôle équivalent dans ces deux systèmes. Chez les Bactéries, l'hydrolyse semble intervenir dans la formation et la stabilisation du complexe protéique (NICHOLS ET AL, 2013), alors que chez les systèmes Archée/Eucaryote, l'hydrolyse de l'ATP n'est pas lié à la stabilité du complexe, mais plutôt à l'interaction avec l'ARNt (Partie II).

Ainsi, alors qu'en première apparence les complexe DEZ/KEOPS semblent fonctionner de la même manière, le rôle respectif de leur activité ATPase semble traduire des mécanismes de fonctionnement différents. **L'étude de ces activités ATPases et la compréhension exacte de leur rôle est donc essentielle pour comprendre les différences qui existent entre les deux systèmes.**

Enfin, une étude similaire à celle entreprise ici doit être réalisée au niveau de la machinerie bactérienne pour comprendre précisément le rôle des partenaires YeaZ et YjeE, et savoir s'ils jouent des rôles équivalents aux protéines Pcc1, Bud32 et Cgi121.

#### **Intérêt pour la recherche appliquée :**

*La compréhension des différences mécanistiques entre la machinerie Eucaryote et la machinerie Bactérienne pourrait permettre, à terme, de trouver des molécules inhibant spécifiquement la machinerie bactérienne. Cette machinerie étant essentielle au fonctionnement de la cellule, cette recherche ouvre la voie à la découverte de nouveaux antibiotiques. Il est donc important de poursuivre la caractérisation détaillée et comparée des protéines constituant ces systèmes respectifs.*

## Des déterminants universels pour la reconnaissance de l'ARNt...

Quel que soit les domaines du vivant, le nucléoside  $t^6A$  n'est retrouvé que sur certains ARNt, portant un anticodon NNU (reconnaissant les codons de la forme ANN)(MODOMICS, MACHNICKA ET AL., 2013). Nous avons montré que les différentes machineries, Eucaryotes, Archées, mitochondriales, peuvent utiliser comme substrats des ARNt issus de différents domaines du vivant, s'ils correspondent effectivement à cette classe d'ARN (NNU). Ainsi, les différentes machineries ont probablement des déterminants généraux communs et universels. Les études sur le Xénope (MORIN ET AL., 1998) ont identifié ces déterminants pour la machinerie Eucaryote, et **il serait donc intéressant, par l'utilisation de divers substrats, de réaliser des études similaires *in vitro* pour déterminer exactement les éléments qui sont reconnus par les machineries au niveau de l'ARNt, et quelles sont les protéines qui en sont responsables**. La protéine Qri7 représente pour cette étude un modèle de choix, car elle ne fait pas appel à des partenaires pour cette reconnaissance. Par comparaison, la compréhension de ces déterminants pourrait donc permettre de mieux comprendre le rôle des partenaires respectifs chez les autres machineries plus complexes.

## ... mais une exception intrigante.

Cependant, malgré une apparente universalité quant aux types de substrats utilisés, une différence fondamentale distingue les différentes machineries : l'ARNt<sup>Met</sup> initiateur. Chez tous les organismes, un ARNt méthionine spécifique, appelé initiateur, est utilisé pour la reconnaissance du premier codon du cadre de lecture. Chez les Eucaryotes, cet ARNt<sup>Meti</sup> ne s'apparie qu'avec un codon start AUG ; alors que chez les Bactéries, les Archées et les organites, l'ARNt<sup>Meti</sup> peut s'apparier avec un codon UUG et CUG dans certains cas (DUBE ET AL. 1968, LECRINIER ET AL., 2006). L'ARNt<sup>Meti</sup>, tout comme l'ARNt<sup>Met</sup> élongateur, possède un anticodon de type NNU, et fait partie de la liste des ARNt susceptibles de porter  $t^6A$ . Chez les Eucaryotes, la règle est effectivement respectée, et ces deux types d'ARNt portent  $t^6A$ . Or chez les Archées, les Bactéries et les Organites, l'ARN<sup>Meti</sup> (à la différence de l'ARN<sup>Meti</sup>) ne porte pas  $t^6A$  ! Cette différence a été invoquée pour expliquer l'initiation alternative chez ces organismes : l'absence de  $t^6A$  autoriserait des appariements moins stricts pour le codon

d'initiation. Cependant, la compréhension de cette différence au niveau de la synthèse de t<sup>6</sup>A reste mystérieuse.

Un premier élément de réponse a été donné il y a maintenant plus de 25 ans (Grosjean et al, 1987). L'ARN<sup>Meti</sup> *E.coli* injecté dans un ovocyte de Xénope est modifié en t<sup>6</sup>A. Cela implique donc que la machinerie de synthèse Eucaryote peut reconnaître cet ARNt comme substrat. Ainsi, l'absence naturelle de t<sup>6</sup>A *in vivo* chez cet ARNt ne provient pas d'une propriété intrinsèque à la molécule en elle-même, mais repose sur des modalités de discrimination de la machinerie bactérienne différente de celle des eucaryotes. A présent que ces machineries sont caractérisées, on peut proposer que ce sont les partenaires YjeE et YeaZ, spécifique aux Bactéries, qui sont responsable de cette sélectivité et permettent de ne pas utiliser l'ARN<sup>Meti</sup> comme substrat. Cependant, une énigme demeure : la machinerie Archéenne est homologue de celle des Eucaryotes. On s'attend donc à ce qu'elle ait des déterminants similaires, et qu'elle reconnaisse donc l'ARNt<sup>Meti</sup> comme substrat, ce qui ne n'est pas le cas *in vivo*.

Des mécanismes de sélectivité différentielle au niveau de cet ARNt pourraient refléter un phénomène d'adaptation spécifique aux « Procaryotes », permettant une plus grande flexibilité d'initiation de la traduction des gènes chez les Archées et les Bactéries, organismes au métabolisme comparable. Il serait intéressant de comprendre également s'il s'agit d'un phénomène d'évolution convergente. A présent que nous avons en main les différentes machineries purifiées, il est donc possible d'étudier la formation de t<sup>6</sup>A *in vitro* en utilisant comme substrats des ARNt<sup>Meti</sup> issus des trois domaines du vivant et de la mitochondrie, afin de comprendre pourquoi et comment cet ARNt particulier fait l'objet d'une sélectivité différentielle. **Cette étude ciblée pourrait ouvrir la voie à la une compréhension plus globale des mécanismes de sélectivité en jeux au sein des différentes machineries.**

#### **Intérêt pour la recherche appliquée :**

*La détermination exacte des déterminants permettant aux machineries de reconnaître l'ARNt substrat est particulièrement important dans le cas de la mitochondrie. Comme nous l'avons vu dans la partie III, de nombreuses maladies mitochondriales sont associées à des mutations dans les ARNt mt. Certaines de ces mutations pourraient affecter l'identité des*

*ARNt, qui ne pourraient plus être reconnu par Qri7. La compréhension des déterminants de cette protéine pour son substrat pourrait donc mener à des découverte de l'origine de pathologies associées à des dysfonctionnements de la mitochondrie.*

### Les composants de la t<sup>6</sup>A synthase, impliqués dans d'autres fonctions biologiques ?

La caractérisation des différentes machineries de synthèse de t<sup>6</sup>A permet maintenant de comprendre la pléiotropie de phénotypes observées chez les différents mutants des protéines composants ces machineries (décrits dans l'introduction). L'absence de t<sup>6</sup>A, causée par une machinerie non fonctionnelle chez ces mutants, est probablement à l'origine de graves erreurs de traduction (EL YACOUBI ET AL, 2011, DAUGERON ET AL 2011) , conduisant à un protéome anormal. Il est probable que seuls certains gènes soient particulièrement affectés par cette absence. **Il est donc maintenant primordial de déterminer ces gènes, pour faire un lien entre chaque phénotype et l'absence de t<sup>6</sup>A.** Nous avons vu dans la partie III que l'étude de Qri7 se révèle un modèle de choix dans cette recherche, grâce au faible nombre de gène candidats mitochondriaux

Cette recherche pourrait également révéler des rôles alternatifs pour les composants de ces machineries. En effet, plusieurs indices suggèrent que la synthèse de t<sup>6</sup>A n'est pas nécessairement la seule fonction biologique de ces protéines.

Chez les Bactéries, le gène codant la protéine RimI est trouvée en association chez de nombreuses espèces aux gènes *yeaZ* et *ygjD*. Dans certains cas, ces gènes sont même fusionnés. Chez ces espèces, la protéine RimI fait donc nécessairement partie du complexe DEZ. Or, cette protéine a comme fonction l'acétylation d'une protéine S18 ribosomale. Sa présence pourrait donc traduire un autre rôle pour le complexe DEZ, lié peut être à la maturation du ribosome. De plus, il a été suggéré qu'YjeE soit impliquée dans la synthèse du peptidoglycane, en raison de son absence chez certaines Mollicutes, qui sont dépourvue de ce composé spécifiquement bactérien (TEPLYAKOV, 2002). Cette hypothèse a été reprise récemment par Oberto et al., qui a montré *in silico* que YeaZ est également absente chez certaines Mollicutes (OBERTO ET AL. 2013). De plus, l'analyse du contexte génomique montre

que le gène codant l'alanine racémase est souvent associé au gène *yjeE*, et parfois même fusionné. Or, l'alanine racémase l'une des enzymes impliquées dans la synthèse du peptidoglycane. YjeE, ou l'ensemble du complexe DEZ, pourrait donc être également orienté vers cette voie métabolique. La détermination de ces autres fonctions pourrait également déboucher sur des hypothèses évolutives intéressantes, puisque le peptidoglycane est spécifique aux bactéries, tout comme YeaZ et YjeE.

Chez les Eucaryotes et les Archées, de même, il n'est pas impossible que les membres du complexe KEOPS puissent être associés à d'autres fonctions. La protéine Bud32 a fait l'objet de nombreuses études mettant en évidence l'existence de différents partenaires ou substrats potentiels. Nous avons mis en évidence que, dans le cadre de KEOPS, elle n'agit pas comme une kinase. Cela n'exclut pas que dans certaines conditions, elle soit effectivement impliquée dans des voies de signalisation cellulaire faisant appel à son activité de phosphotransférase. Enfin, la recherche de nouveaux partenaires aux protéines du complexe KEOPS pourrait également révéler des mécanismes de régulation de son activité. Récemment, il a été chez l'Homme qu'une ubiquitine-ligase E3 est recrutée au niveau du complexe KEOPS (COSTESSI ET AL, 2012). De plus, Kae1, Bud32 et Cgi121 ont été identifiés par étude protéomique comme étant effectivement ubiquitinylée chez l'Homme (KIM ET AL., 2011). De manière intéressante, chez certaines Archées, le gène Cgl121 est retrouvé fusionné avec une protéine appelée Moad, qui joue le même rôle que l'ubiquitine chez les Eucaryotes (MAKAROVA ET AL, 2010). L'étude du lien entre KEOPS et l'ubiquitinylation pourrait à terme révéler des processus de régulation conservés entre les Archées et les Eucaryotes. Nous avons commencé à explorer cette voie, en purifiant les deux orthologues de Moad présent chez *Pyrococcus abyssi* et leur activateur ubiquitine-ligase like, MoeB (DONNEES NON PUBLIES, TRAVAUX EN COURS !).

**De nombreuses pistes peuvent donc encore être explorées, en parallèle à la poursuite de la caractérisation de Kae1, Bud32, Cgi121, Pcc1, Gon7, YgjD, YeaZ, YjeE, YrdC, Sua5, Qri7 dans la voie de biosynthèse de t<sup>6</sup>A, qui ne manqueront pas de révéler des surprises inattendues, car comme le dit J.R.R Tolkien (dans *Bilbo le Hobbit*) « Quand on cherche, on trouve généralement quelque chose, mais ce n'est pas toujours exactement ce qu'on voulait ».**



## Epilogue : une histoire de nom...

*«Il y a bien longtemps, la déesse Isis, femme du dieu Osiris, était réputée pour être une redoutable magicienne de l’Egypte Antique. Pourtant, à cette époque, le Dieu le plus puissant était Râ, dieu solaire et père du panthéon égyptien. Il était dit que Râ possédait un nom secret qui renfermait tous ses pouvoirs. Pour asseoir son fils Horus à la tête du panthéon, Isis mis en place un stratagème malicieux. Un jour où Râ était assoupi, elle récolta un peu de salive du dieu et en façonna un serpent qu’elle dissimula dans les buissons. A son réveil, le dieu Râ fut mordu par ce serpent et fut pris d’une douleur insupportable. Seule Isis détenait l’antidote pouvant délivrer le dieu Râ du venin qui le torturait. Elle lui proposa de le guérir, à condition que le dieu lui révèle son nom secret. A l’agonie, Râ fut ainsi contraint de répondre à la requête d’Isis : la déesse, connaissant alors ce nom, devint la plus puissante des magiciennes »*

Ce mythe illustre une notion fondamentale pour les anciens égyptiens : le pouvoir des noms. Selon eux, un nom pouvait donner la vie, ou la reprendre. Aujourd’hui encore, les noms ont une symbolique puissante, notamment dans le domaine de la biologie. Nommer une protéine, c’est lui assigner une fonction : nomenclature et rôle biologique sont donc intimement liés. De 2009 à 2013, les très nombreuses publications concernant la voie de biosynthèse de t<sup>6</sup>A ont permis de révéler la fonction de différentes protéines, dont le nom ne reflétait que certaines de leurs caractéristiques biologiques: il est donc nécessaire de réfléchir à nouveau à la nomenclature de ces protéines.

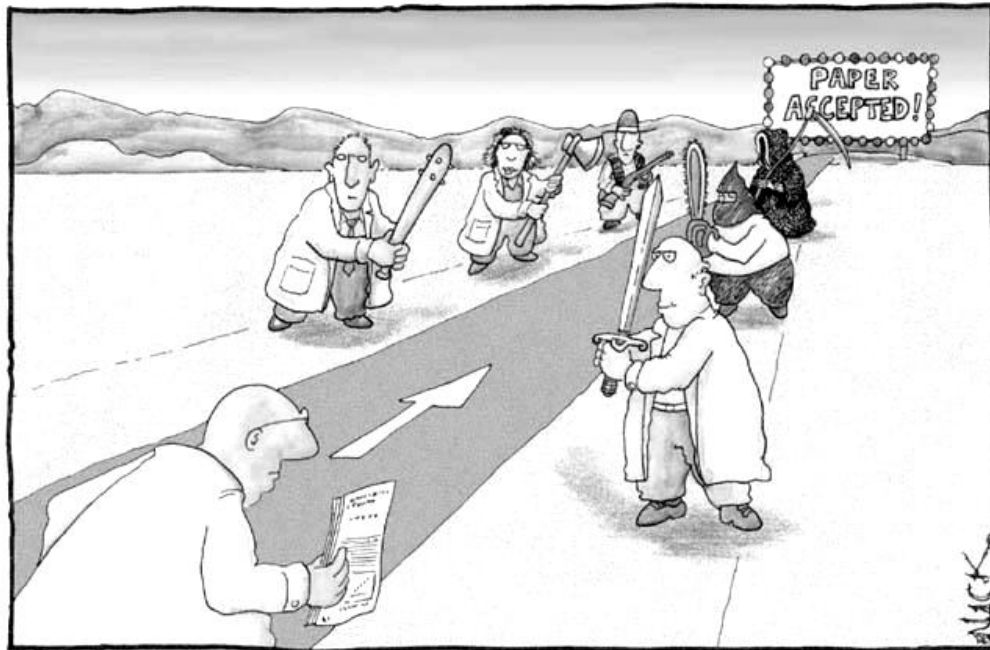
Chez les Bactéries, ce travail a déjà été entrepris : l’équipe de Valérie De Crécy-Lagard a proposé de remplacer les anciennes nomenclatures **YxxX** (Y signifiant fonction inconnue) par **TsaX** (pour t-six-A) en conservant une part d’historique : YrdC devient TsaC, YgjD devient TsaD, YjeE devient TsaE et YeaZ devient TsaB (suivant l’alphabet), TsaA étant l’enzyme responsable de la modification de t<sup>6</sup>A en mt<sup>6</sup>A. Ces noms ont l’avantage d’indiquer clairement la fonction de ces protéines, et apportent aussi une cohérence. Cependant, ils peuvent laisser penser que ces protéines sont exprimées en opéron, ou qu’elles forment une réaction linéaire  $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ .

Et chez les Archées et les Eucaryotes ? Le nom **Kae1** (kinase-associated endopeptidase) est ici totalement obsolète : ce n'est pas une endopeptidase, mais une transférase, et elle est associée à une ATPase plutôt qu'à une kinase. Il serait envisageable de la renommer **e/aTsaD** (pour ekarya/archaea), soulignant l'orthologie avec **TsaD**. **Qri7** pourrait ainsi être appelée **mtTsaD** (pour mitochondrial). Mais qu'en est-il de Bud32 ou PRPK chez l'Homme ; Cgi121 et Pcc1 ? En suivant la même logique que pour les Bactéries, il serait possible de les renommer : **TsaF**, **TsaG**, **TsaH**. Mais cette uniformisation des nomenclatures **TsaX** ne risquerait-elle pas de faire oublier une information essentielle, à savoir les différences entre le système Bactérien et Archée/Eucaryote ? De plus, le nom *Tsa* exclut inévitablement l'existence possible d'autres fonctions pour ces protéines. La nomenclature de Sua5 est également complexe : ne faudrait-il pas renommer tous les orthologues **TsaC** (possédant tous le domaine « YrdC ») et ajouter un **S** pour préciser la présence d'un domaine SUA5 supplémentaire dans certains cas ? (chez *P.abyssi* et *S.cerevisiae*, la protéine serait donc **TsaC-S**). Au bilan, le problème de nomenclature semble être soumis à des paramètres complexes, qui rendent difficile un changement immédiat.

Enfin, qu'en est-il du complexe **KEOPS**, dont le nom est également obsolète au vue des travaux présentés ici (Kinase, Endopeptidase and Other Proteins of Small size) ? Un tel complexe pourrait éventuellement être renommé pour refléter sa fonction ou son activité, puisque nous avons vu que chacun de ses composants participe à la synthèse de t<sup>6</sup>A. Je propose ainsi le nom de **ATON** : AMP-thréonylcarbamoyl Transférase On N<sup>6</sup>-adenosine, qui refléterait l'activité biochimique du complexe. Historiquement, Aton est un dieu solaire unique instauré par le pharaon Akhénaton (père du célèbre Toutankhamon et époux de Nefertiti) au XIV siècle av. JC. Cependant, à sa mort, ce culte d'Aton fut considéré comme une hérésie, et le nom de Aton et du pharaon associé furent systématiquement effacés de tous les temples. Cette nouvelle proposition est-elle vouée au même destin ? En effet, comme toutes les nomenclatures, on peut lui trouver ses défauts...

Une autre appellation est aussi possible : **ATTOR** ; pour ATPase, Threonylcarbamoyl Transferase and Other Regulatory proteins. Ici, la logique de la nomenclature d'origine de KEOPS est conservée, en remplaçant simplement l'activité putative des protéines par leur activité démontrée dans ces travaux. Il est à noter que Hathor était la déesse égyptienne de la fécondité, symbole de fertilité... et de chance !

# BIBLIOGRAPHIE



Most scientists regarded the new streamlined peer-review process as 'quite an improvement.'

---



**A**bdullah KM, Lo RY, Mellors A. Cloning, nucleotide sequence, and expression of the *Pasteurella haemolytica* A1 glycoprotease gene. J Bacteriol. 1991 Sep;173(18):5597-603.

Abe Y, Matsumoto S, Wei S, Nezu K, Miyoshi A, Kito K, Ueda N, Shigemoto K, Hitsumoto Y, Nikawa J, Enomoto Y. Cloning and characterization of a p53-related protein kinase expressed in interleukin-2-activated cytotoxic T-cells, epithelial tumor cell lines, and the testes. J Biol Chem. 2001 Nov 23;276(47):44003-11

Abe Y, Takeuchi T, Imai Y, Murase R, Kamei Y, Fujibuchi T, Matsumoto S, Ueda N, Ogasawara M, Shigemoto K, Kito K. A Small Ras-like protein Ray/Rab1c modulates the p53-regulating activity of PRPK. Biochem Biophys Res Commun. 2006 May 26;344(1):377-85.

Agari Y, Sato S, Wakamatsu T, Bessho Y, Ebihara A, Yokoyama S, Kuramitsu S, Shinkai A. X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. Proteins. 2008 Feb 15;70(3):1108-11

Agari Y, Sato S, Wakamatsu T, Bessho Y, Ebihara A, Yokoyama S, Kuramitsu S, Shinkai A. X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. Proteins. 2008 Feb 15;70(3):1108-11.

Agris PF, Vendeix FA, Graham WD. tRNA's wobble decoding of the genome: 40 years of modification. J Mol Biol. 2007 Feb 9;366(1):1-13..

Agris PF. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. EMBO Rep. 2008 Jul;9(7):629-35.

Albers SV, Meyer BH. The archaeal cell envelope. Nat Rev Microbiol. 2011 Jun;9(6):414-26..

Allali-Hassani A, Campbell TL, Ho A, Schertzer JW, Brown ED. Probing the active site of YjeE: a vital *Escherichia coli* protein of unknown function. Biochem J. 2004 Dec 15;384(Pt 3):577-84.

Andersson SG, Karlberg O, Canbäck B, Kurland CG. On the origin of mitochondria: a genomics perspective. Philos Trans R Soc Lond B Biol Sci. 2003 Jan 29;358(1429):165-77; discussion 177-9. Review..

Anton BP, Russell SP, Vertrees J, Kasif S, Raleigh EA, Limbach PA, Roberts RJ. Functional characterization of the YmcB and YqeV tRNA methyltransferases of *Bacillus subtilis*. Nucleic Acids Res. 2010 Oct;38(18):6195-205

Aravind L, Koonin EV. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol. 1999 Apr 16;287(5):1023-40.

Arcari P, Brownlee GG. Nucleic Acids Res. 1980 Nov 25;8(22):5207-12. The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria.

Arcari P, Brownlee GG. The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria. Nucleic Acids Res. 1980 Nov 25;8(22):5207-12.

Arigoni F, Talabot F, Peitsch M, Edgerton MD, Meldrum E, Allet E, Fish R, Jamotte T, Curchod ML, Loferer H. A genome-based approach for the identification of essential bacterial genes. Nat Biotechnol. 1998 Sep;16(9):851-6.

Arigoni F, Talabot F, Peitsch M, Edgerton MD, Meldrum E, Allet E, Fish R, Jamotte T, Curchod ML, Loferer H. A genome-based approach for the identification of essential bacterial genes. Nat Biotechnol. 1998 Sep;16(9):851-6.

Arragain S, Handelman SK, Forouhar F, Wei FY, Tomizawa K, Hunt JF, Douki T, Fontecave M, Mulliez E, Atta M. Identification of eukaryotic and prokaryotic methyltransferase for biosynthesis of 2-methylthio-N<sup>6</sup>-threonylcarbamoyladenine in tRNA. J Biol Chem. 2010 Sep 10;285(37):28425-33.

Atkins JF, Gesteland R. Biochemistry. The 22nd amino acid. Science. 2002 May 24;296(5572):1409-10

Aydin I, Saijo-Hamano Y, Namba K, Thomas C, Roujeinikova A. Structural analysis of the essential resuscitation promoting factor YeaZ suggests a mechanism of nucleotide regulation through dimer reorganization. PLoS One. 2011;6(8):e23245.

**B**ajji AC, Davis DR. Synthesis of the tRNA(Lys,3) anticodon stem-loop domain containing the hypermodified ms2t6A nucleoside. J Org Chem. 2002 Jul 26;67(15):5352-8.

Benelli D, Londei P. Begin at the beginning: evolution of translational initiation. Res Microbiol. 2009 Sep;160(7):493-501.

Ben-Menachem R, Tal M, Shadur T, Pines O. A third of the yeast mitochondrial proteome is dual localized: a question of evolution. Proteomics. 2011 Dec;11(23):4468-76.

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition) New York: W H Freeman; 2002.

Björk GR. Prog Nucleic Acid Res Mol Biol. 1995;50:263-338. Review Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA.

**Blöchl E, Rachel R, Burggraf S, Hafenbradl D, Jannasch HW, Stetter KO.** *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles*. 1997 Feb;1(1):14-21.

**Boguta M, Hunter LA, Shen WC, Gillman EC, Martin NC, Hopper AK.** Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol. *Mol Cell Biol*. 1994 Apr;14(4):2298-306.

**Bory de Saint-Vincent, J.B.G.M** - Règne psychodiaré. Article extrait du tome II (Zoophytes), histoire naturelle, de l'"Encyclopédie méthodique", et faisant suite à l'"Essai sur les microscopiques"

**Boucher Y, Kamekura M, Doolittle WF.** Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol*. 2004 Apr;52(2):515-27.

**Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P.** Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol*. 2008 Mar;6(3):245-52.

**Brykailo MA, Corbett AH, Fridovich-Keil JL.** Functional overlap between conserved and diverged KH domains in *Saccharomyces cerevisiae* SCP160. *Nucleic Acids Res*. 2007;35(4):1108-18.

**Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A.** Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*. 2005 Feb 3;433(7025):531-7.

**Cavicchioli R.**, Textbook "Archaea : Molecular and Cellular Biology", 2007, ASM Press, Washington, D.C.

**Chatton É.**, *Titres et Travaux Scientifiques*, Sottano (Sète, Italie), 1937

**Chen J, Ji C, Gu S, Zhao E, Dai J, Huang L, Qian J, Ying K, Xie Y, Mao Y.** Isolation and identification of a novel cDNA that encodes human yrdC protein. *J Hum Genet*. 2003;48(4):164-9.

**Chheda GB, Hall RH, Mozejko J, Magrath DI, Schweizer MP, Stasiuk L, Taylor PR.** Aminoacyl nucleosides. VI. Isolation and preliminary characterization of threonyladenine derivatives from transfer ribonucleic acid. *Biochemistry*. 1969 Aug;8(8):3278-82

**Christian T, Evilia C, Williams S, Hou YM.** Distinct origins of tRNA(m1G37) methyltransferase. *J Mol Biol*. 2004 Jun 11;339(4):707-19.

**Christian T, Hou YM.** Distinct determinants of tRNA recognition by the TrmD and Trm5 methyl transferases. *J Mol Biol*. 2007 Oct 26;373(3):623-32.

**Clemente ML, Sartori G, Cardazzo B, Carignani G.** Analysis of an 11.6 kb region from the right arm of chromosome VII of *Saccharomyces cerevisiae* between the RAD2 and the MES1 genes reveals the presence of three new genes. *Yeast*. 1997 Mar 15;13(3):287-90.

**Cochella L, Brunelle JL, Green R.** Mutational analysis reveals two independent molecular requirements during transfer RNA selection on the ribosome. *Nat Struct Mol Biol*. 2007 Jan;14(1):30-6.

**Coque JJ, Pérez-Llarena FJ, Enguita FJ, Fuente JL, Martín JF, Liras P.** Characterization of the cmcH genes of *Nocardia lactamdurans* and *Streptomyces clavuligerus* encoding a functional 3'-hydroxymethylcephem O-carbamoyltransferase for cephamycin biosynthesis. *Gene*. 1995 Aug 30;162(1):21-7.

**Costessi A, Mahrour N, Sharma V, Stunnenberg R, Stoel MA, Tijchon E, Conaway JW, Conaway RC, Stunnenberg HG.** The human EKC/KEOPS complex is recruited to Cullin2 ubiquitin ligases by the human tumour antigen PRAME. *PLoS One*. 2012;7(8):e42822.

**Crick FH.** Codon-anticodon pairing: the wobble hypothesis. *J Mol Biol*. 1966 Aug;19(2):548-55

**Crick FH.** The origin of the genetic code. *J Mol Biol*. 1968 Dec;38(3):367-79.

**Cunningham RS, Gray MW.** Derivatives of N-(N-(9-beta-D-ribofuranosylpurin-6-yl)carbamoyl)threonine in phosphodiesterase hydrolysates of wheat embryo transfer ribonucleic acid. *Biochemistry*. 1974 Jan 29;13(3):543-53.

**Dalluge JJ, Hamamoto T, Horikoshi K, Morita RY, Stetter KO, McCloskey JA.** Posttranscriptional modification of tRNA in psychrophilic bacteria. *J Bacteriol*. 1997 Mar;179(6):1918-23.

**Daugeron MC, Lenstra TL, Frizzarin M, El Yacoubi B, Liu X, Baudin-Baillieu A, Lijnzaad P, Decourty L, Saveanu C, Jacquier A, Holstege FC, de Crécy-Lagard V, van Tilbeurgh H, Libri D.** Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t6A modification of tRNAs. *Nucleic Acids Res*. 2011 Aug;39(14):6148-60.

**Davis FF, Allen FW.** Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem*. 1957 Aug;227(2):907-15

**DeLong EF, Pace NR.** Syst Biol. 2001 Aug;50(4):470-8. Environmental diversity of bacteria and archaea.

**Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S.** *On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes.* Res Microbiol. 2011 Jan;162(1):53-70

**Deutsch C, El Yacoubi B, de Crécy-Lagard V, Iwata-Reuyl D.** *Biosynthesis of threonylcarbamoyl adenosine (t6A), a universal tRNA nucleoside.* J Biol Chem. 2012 Apr 20;287(17):13666-73.

**D'Haeze W, Van Montagu M, Promé JC, Holsters M.** *Carbamoylation of azorhizobial Nod factors is mediated by NodU.* Mol Plant Microbe Interact. 1999 Jan;12(1):68-73.

**Dias CL, Ala-Nissila T, Wong-ekkabut J, Vattulainen I, Grant M, Karttunen M.** *The hydrophobic effect and its role in cold denaturation.* Cryobiology. 2010 Feb;60(1):91-9.

**Dihanich ME, Najarian D, Clark R, Gillman EC, Martin NC, Hopper AK.** *Isolation and characterization of MOD5, a gene required for isopentenylolation of cytoplasmic and mitochondrial tRNAs of Saccharomyces cerevisiae.* Mol Cell Biol. 1987 Jan;7(1):177-84.

**Downey M, Houlsworth R, Maringele L, Rollie A, Brehme M, Galicia S, Guillard S, Partington M, Zubko MK, Krogan NJ, Emili A, Greenblatt JF, Harrington L, Lydall D, Durocher D.** *A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator.* Cell. 2006 Mar 24;124(6):1155-68.

**Downey M, Houlsworth R, Maringele L, Rollie A, Brehme M, Galicia S, Guillard S, Partington M, Zubko MK, Krogan NJ, Emili A, Greenblatt JF, Harrington L, Lydall D, Durocher D.** *A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator.* Cell. 2006 Mar 24;124(6):1155-68.

**Dube SK, Marcker KA, Clark BF, Cory S.** *Nucleotide sequence of N-formyl-methionyl-transfer RNA.* Nature. 1968 Apr 20;218(5138):232-3.

**Durant PC, Bajji AC, Sundaram M, Kumar RK, Davis DR.** *Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNA<sup>Lys</sup> anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A.* Biochemistry. 2005 Jun 7;44(22):8078-89.

**El Yacoubi B, Hatin I, Deutsch C, Kahveci T, Rousset JP, Iwata-Reuyl D, Murzin AG, de Crécy-Lagard V.** *A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification.* EMBO J. 2011 Mar 2;30(5):882-93.

**El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, de Crécy-Lagard V.** *The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA.* Nucleic Acids Res. 2009 May;37(9):2894-909.

**Elkins BN, Keller EB.** *The enzymatic synthesis of N-(purin-6-ylcarbamoyl)threonine, an anticodon-adjacent base in transfer ribonucleic acid.* Biochemistry. 1974 Oct 22;13(22):4622-8.

**Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin EV, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO.** *A korarchaeal genome reveals insights into the evolution of the Archaea.* Proc Natl Acad Sci U S A. 2008 Jun 10;105(23):8102-7..

**Erauso, G., F. Charbonnier, T. Barbeyron, P. Forterre, and D. Prieur. 1992.** *Preliminary characterization of an hyperthermophilic archaeon bearing a plasmid isolated from a North Fijian basin hydrothermal vent.* C. R. Acad. Sci. 314:387–393.

**Eriani G, Cavarelli J, Martin F, Dirheimer G, Moras D, Gangloff J.** *Role of dimerization in yeast aspartyl-tRNA synthetase and importance of the class II invariant proline.* Proc Natl Acad Sci U S A. 1993 Nov 15;90(22):10816-20.

**Facchin S, Lopreiato R, Ruzzene M, Marin O, Sartori G, Götz C, Montenarh M, Carignani G, Pinna LA.** *Functional homology between yeast piD261/Bud32 and human PRPK: both phosphorylate p53 and PRPK partially complements piD261/Bud32 deficiency.* FEBS Lett. 2003 Aug 14;549(1-3):63-6.

**Facchin S, Lopreiato R, Stocchetto S, Arrigoni G, Cesaro L, Marin O, Carignani G, Pinna LA.** *Structure-function analysis of yeast piD261/Bud32, an atypical protein kinase essential for normal cell life.* Biochem J. 2002 Jun 1;364(Pt 2):457-63.

**Facchin S, Ruzzene M, Peggion C, Sartori G, Carignani G, Marin O, Brustolon F, Lopreiato R, Pinna LA.** *Phosphorylation and activation of the atypical kinase p53-related protein kinase (PRPK) by Akt/PKB.* Cell Mol Life Sci. 2007 Oct;64(19-20):2680-9.

**Facchin S, Sarno S, Marin O, Lopreiato R, Sartori G, Pinna LA.** *Acidophilic character of yeast PID261/BUD32, a putative ancestor of eukaryotic protein kinases.* Biochem Biophys Res Commun. 2002 Sep 6;296(5):1366-71.

**Ferreira-Cerca S, Sagar V, Schäfer T, Diop M, Wesseling AM, Lu H, Chai E, Hurt E, LaRonde-LeBlanc N.** *ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit.* Nat Struct Mol Biol. 2012 Dec;19(12):1316-23.

**Files JG, Weber K, Coulondre C, Miller JH.** *Identification of the UUG codon as a translational initiation codon in vivo.* J Mol Biol. 1975 Jun 25;95(2):327-30.

**Forterre P, Gribaldo S, Brochier C.** *[Luca: the last universal common ancestor].* Med Sci (Paris). 2005 Oct;21(10):860-5. Review. French.

**Forterre P.** *A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein.* Trends Genet. 2002 May;18(5):236-7.

**Forterre P.** *Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain.* Proc Natl Acad Sci U S A. 2006 Mar 7;103(10):3669-74.

**Foury F, Roganti T, Lecrenier N, Purnelle B.** *The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae.* FEBS Lett. 1998 Dec 4;440(3):325-31.

**Fox GE.** *Origin and evolution of the ribosome.* Cold Spring Harb Perspect Biol. 2010 Sep;2(9):a003483.

**Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA 3rd, Venter JC.** *The minimal gene complement of Mycoplasma genitalium.* Science. 1995 Oct 20;270(5235):397-403.

**Freiberg C, Wieland B, Spaltmann F, Ehlert K, Brötz H, Labischinski H.** *Identification of novel essential Escherichia coli genes conserved among pathogenic bacteria.* J Mol Microbiol Biotechnol. 2001 Jul;3(3):483-9.

**Galperin MY, Koonin EV.** *'Conserved hypothetical' proteins: prioritization of targets for experimental study.* Nucleic Acids Res. 2004 Oct 12;32(18):5452-63..

**Galperin MY, Koonin EV.** *From complete genome sequence to 'complete' understanding?* Trends Biotechnol. 2010 Aug;28(8):398-406.

**Gerdes S, El Yacoubi B, Bailly M, Blaby IK, Blaby-Haas CE, Jeanguenin L, Lara-Núñez A, Pribat A, Waller JC, Wilke A, Overbeek R, Hanson AD, de Crécy-Lagard V.** *Synergistic use of plant-prokaryote comparative genomics for functional annotations.* BMC Genomics. 2011 Jun 15;12 Suppl 1:S2

**Gerhard Fiala, Karl O. Stetter** *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C Archives of Microbiology June 1986, Volume 145, Issue 1, pp 56-61

**Germot A, Philippe H, Le Guyader H.** *Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in Nosema locustae.* Mol Biochem Parasitol. 1997 Aug;87(2):159-68.

**Gil R, Silva FJ, Peretó J, Moya A.** *Determination of the core of a minimal bacterial gene set.* Microbiol Mol Biol Rev. 2004 Sep;68(3):518-37, table

**Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ.** *Genome streamlining in a cosmopolitan oceanic bacterium.* Science. 2005 Aug 19;309(5738):1242-5.

**Goto-Ito S, Ito T, Ishii R, Muto Y, Bessho Y, Yokoyama S.** *Crystal structure of archaeal tRNA(m(1)G37)methyltransferase aTrm5.* Proteins. 2008 Sep;72(4):1274-89..

**Grant J, Han GW, Jin KK, Klock HE, Knuth MW, Miller MD, Morse AT, Nigoghossian E, Okach L, Oommachen S, Paulsen J, Reyes R, Rife CL, van den Bedem H,**

**Grosjean H, de Crécy-Lagard V, Marck C.** *Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.* FEBS Lett. 2010 Jan 21;584(2):252-64

**Grosjean H, De Henau S, Doi T, Yamane A, Ohtsuka E, Ikehara M, Beauchemin N, Nicoghossian K, Cedergren R.** *The in vivo stability, maturation and aminoacylation of anticodon-substituted Escherichia coli initiator methionine tRNAs.* Eur J Biochem. 1987 Jul 15;166(2):325-32.

**Grosjean H, Marck C, de Crécy-Lagard V.** *The various strategies of codon decoding in organisms of the three domains of life: evolutionary implications.* Nucleic Acids Symp Ser (Oxf). 2007;(51):15-6.

**Grosjean H, Sprinzl M, Steinberg S.** *Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies.* Biochimie. 1995;77(1-2):139-41.

**Grosjean H., Benne R.,** « *Modification and RNA editing of RNA* » ASM Press, 1998, Textbook

**Gustilo EM, Vendeix FA, Agris PF.** *tRNA's modifications bring order to gene expression.* Curr Opin Microbiol. 2008 Apr;11(2):134-40.

**Haeckel, E. H. P. A.** *Generelle Morphologie der Organismen : allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von C. Darwin reformirte Decendenz-Theorie.* 1866.Berlin.

**Hampsey M, Na JG, Pinto I, Ware DE, Berroteran RW.** *Extragenic suppressors of a translation initiation defect in the cyc1 gene of Saccharomyces cerevisiae.* Biochimie. 1991 Dec;73(12):1445-55.



**Handford JI, Ize B, Buchanan G, Butland GP, Greenblatt J, Emili A, Palmer T.** *Conserved network of proteins essential for bacterial viability.* J Bacteriol. 2009 Aug;191(15):4732-49.

**Handford JI, Ize B, Buchanan G, Butland GP, Greenblatt J, Emili A, Palmer T.** *Conserved network of proteins essential for bacterial viability.* J Bacteriol. 2009 Aug;191(15):4732-49.

**Harada F, Matsubara M, Kato N.** *Stable tRNA precursors in HeLa cells.* Nucleic Acids Res. 1984 Dec 21;12(24):9263-9.

**Harris JK, Kelley ST, Spiegelman GB, Pace NR.** *The genetic core of the universal ancestor.* Genome Res. 2003 Mar;13(3):407-12.

**Harris KA, Jones V, Bilbille Y, Swairjo MA, Agris PF.** *YrdC exhibits properties expected of a subunit for a tRNA threonylcarbamoyl transferase.* RNA. 2011 Sep;17(9):1678-87.

**Hausuehl K, Huesgen PF, Meier M, Dessi P, Glaser E, Adamski J, Adamska I.** *Eukaryotic GCP1 is a conserved mitochondrial protein required for progression of embryo development beyond the globular stage in Arabidopsis thaliana.* Biochem J. 2009 Oct 12;423(3):333-41.

**Hecker A, Graille M, Madec E, Gadelle D, Le Cam E, van Tilbergh H, Forterre P.** *The universal Kae1 protein and the associated Bud32 kinase (PRPK), a mysterious protein couple probably essential for genome maintenance in Archaea and Eukarya.* Biochem Soc Trans. 2009 Feb;37(Pt 1):29-35. Erratum in: Biochem Soc Trans. 2009 Apr;37(Pt 2):477.

**Hecker A, Leulliot N, Gadelle D, Graille M, Justome A, Dorlet P, Brochier C, Quevillon-Cheruel S, Le Cam E, van Tilbeurgh H, Forterre P.** *An archaeal orthologue of the universal protein Kae1 is an iron metalloprotein which exhibits atypical DNA-binding properties and apurinic-endonuclease activity in vitro.* Nucleic Acids Res. 2007;35(18):6042-51.

**Hecker A, Lopreiato R, Graille M, Collinet B, Forterre P, Libri D, van Tilbeurgh H.** *Structure of the archaeal Kae1/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex.* EMBO J. 2008 Sep 3;27(17):2340-51.

**Helm M, Giegé R, Florentz C.** *A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNA<sup>Lys</sup>.* Biochemistry. 1999 Oct 5;38(40):13338-46.

**Hodgson KO, Wooley J, Deacon AM, Godzik A, Lesley SA, Wilson IA.** *Structure of an essential bacterial protein YeaZ (TM0874) from Thermotoga maritima at 2.5 Å resolution.* Acta Crystallogr Sect F Struct Biol Cryst Commun. 2010 Oct 1;66(Pt 10):1230-6.

**Howell N.** *Human mitochondrial diseases: answering questions and questioning answers.* Int Rev Cytol. 1999;186:49-116.

**Hunt A, Rawlins JP, Thomaidis HB, Errington J.** *Functional analysis of 11 putative essential genes in Bacillus subtilis.* Microbiology. 2006 Oct;152(Pt 10):2895-907.

**Ishikura H, Yamada Y, Murao K, Saneyoshi M, Nishimura S.** *The presence of N-[9-(beta-D-ribofuranosyl)purin-6-ylcarbamoyl]threonine in serine, methionine and lysine transfer RNA's from Escherichia coli.* Biochem Biophys Res Commun. 1969 Dec 4;37(6):990-5.

**Jarrell KF, Albers SV.** *The archaeellum: an old motility structure with a new name.* Trends Microbiol. 2012 Jul;20(7):307-12.

**Jeudy S, Stelter M, Coutard B, Kahn R, Abergel C.** *Preliminary crystallographic analysis of the Escherichia coli YeaZ protein using the anomalous signal of a gadolinium derivative.* Acta Crystallogr Sect F Struct Biol Cryst Commun. 2005 Sep 1;61(Pt 9):848-51.

**Jia J, Lunin VV, Sauv   V, Huang LW, Matte A, Cygler M.** *Crystal structure of the YciO protein from Escherichia coli.* Proteins. 2002 Oct 1;49(1):139-41.

**Jiang W, Prokopenko O, Wong L, Inouye M, Mirochnitchenko O.** *IRIP, a new ischemia/reperfusion-inducible protein that participates in the regulation of transporter activity.* Mol Cell Biol. 2005 Aug;25(15):6496-508.

**Kaczanowska M, Ryd  n-Aulin M.** *The YrdC protein--a putative ribosome maturation factor.* Biochim Biophys Acta. 2005 Feb 14;1727(2):87-96.

**Karlberg O, Canb  ck B, Kurland CG, Andersson SG.** *The dual origin of the yeast mitochondrial proteome.* Yeast. 2000 Sep 30;17(3):170-87.

**Khudiyakov YuE, Neplyueva VS, Kalinina TI, Smirnov VD.** *Effect of structure of the initiator codon on translation in E. coli.* FEBS Lett. 1988 May 23;232(2):369-71.

**Kilpatrick MW, Walker RT.** *The nucleotide sequence of the tRNA<sup>Met</sup> from the archaeobacterium Thermoplasma acidophilum.* Nucleic Acids Res. 1981 Sep 11;9(17):4387-90.

**Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, Harper JW, Gygi SP.** *8. Systematic and quantitative assessment of the ubiquitin-modified proteome.* Mol Cell. 2011 Oct 21;44(2):325-40.

**Kimura-Harada F, Von Minden DL, McCloskey JA, Nishimura S.** *N-((9-*D*-ribofuranosylpurin-6-yl)-*N*-methylcarbamoyl) threonine, a modified nucleoside isolated from Escherichia coli threonine transfer ribonucleic acid.* Biochemistry. 1972 Oct 10;11(21):3910-5.

**Kisseleva-Romanova E, Lopreiato R, Baudin-Baillieu A, Rousselle JC, Ilan L, Hofmann K, Namane A, Mann C, Libri D.** *Yeast homolog of a cancer-testis antigen defines a new transcription complex.* EMBO J. 2006 Aug 9;25(15):3576-85.

**Koonin EV.** *Comparative genomics, minimal gene-sets and the last universal common ancestor.* Nat Rev Microbiol. 2003 Nov;1(2):127-36.

**Koonin EV.** *Orthologs, paralogs, and evolutionary genomics.* Annu Rev Genet. 2005;39:309-38. Review.

**Körner A, Söll D.** *N-(purin-6-ylcarbamoyl)threonine: biosynthesis in vitro in transfer RNA by an enzyme purified from Escherichia coli.* FEBS Lett. 1974 Mar 1;39(3):301-6.

**Koski RA, Clarkson SG.** *Synthesis and maturation of Xenopus laevis methionine tRNA gene transcripts in homologous cell-free extracts.* J Biol Chem. 1982 Apr 25;257(8):4514-21.

**Kowalak JA, Dalluge JJ, McCloskey JA, Stetter KO.** *The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles.* Biochemistry. 1994 Jun 28;33(25):7869-76.

**Kramer E.B., P.J.** *The frequency of translational misreading errors in E. coli is largely determined by tRNA competition RNA.* 2007 January; 13(1): 87–96.

**Kramer E.B., Vallabhaneni H., Mayer L. M., Farabaugh P. J.** *A comprehensive analysis of translational missense errors in the yeast Saccharomyces cerevisiae.* RNA. 2010 September; 16(9): 1797–1808.

**Kuratani M, Kasai T, Akasaka R, Higashijima K, Terada T, Kigawa T, Shinkai A, Bessho Y, Yokoyama S.** *Crystal structure of Sulfolobus tokodaii Sua5 complexed with L-threonine and AMPPNP.* Proteins. 2011 Jul;79(7):2065-75.

**Kuratani M, Kasai T, Akasaka R, Higashijima K, Terada T, Kigawa T, Shinkai A, Bessho Y, Yokoyama S.** *Crystal structure of Sulfolobus tokodaii Sua5 complexed with L-threonine and AMPPNP.* Proteins. 2011 Jul;79(7):2065-75.

**Lahoud G, Goto-Ito S, Yoshida K, Ito T, Yokoyama S, Hou YM.** *Differentiating analogous tRNA methyltransferases by fragments of the methyl donor.* RNA. 2011 Jul;17(7):1236-46.

**Lamarck, J. B. P. A. de M. de.,** *Histoire naturelle des Animaux sans Vertèbres*, troisième édition, 1837

**Lamichhane TN, Blewett NH, Maraia RJ.** *Plasticity and diversity of tRNA anticodon determinants of substrate recognition by eukaryotic A37 isopentenyltransferases.* RNA. 2011 Oct;17(10):1846-57.

**Lee C, Kramer G, Graham DE, Appling DR.** *Yeast mitochondrial initiator tRNA is methylated at guanosine 37 by the Trm5-encoded tRNA (guanine-N1-)-methyltransferase.* J Biol Chem. 2007 Sep 21;282(38):27744-53.

**Lei T, Yang J, Zheng L, Markowski T, Witthuhn BA, Ji Y.** *The essentiality of staphylococcal Gcp is independent of its repression of branched-chain amino acids biosynthesis.* PLoS One. 2012;7(10):e46836.

**Leipe DD, Koonin EV, Aravind L.** *Evolution and classification of P-loop kinases and related proteins.* J Mol Biol. 2003 Oct 31;333(4):781-815.

**Leonard CJ, Aravind L, Koonin EV.** *Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily.* Genome Res. 1998 Oct;8(10):1038-47.

**Lescliner E, Nauwelaerts K, Zanier K, Poesen K, Sattler M, Herdewijn P.** *The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif.* Nucleic Acids Res. 2006 May 31;34(10):2878-86. Print 2006.

**Linnæi Caroli** (Carl Linné): *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.* Editio decima reformata 1758, Holmiæ, Impensis direct. Laurentii Salvii

**Londei, P.** «Chapter 8 – Translation ». *Textbook Archaea : Molecular and Cellular Biology*, R.Cavicchioli, 2007, ASM Press, Washington, D.C.

**Lopreiato R, Facchin S, Sartori G, Arrigoni G, Casonato S, Ruzzene M, Pinna LA, Carignani G.** *Analysis of the interaction between piD261/Bud32, an evolutionarily conserved protein kinase of Saccharomyces cerevisiae, and the Grx4 glutaredoxin.* Biochem J. 2004 Jan 15;377(Pt 2):395-405.

**Lopreiato R, Facchin S, Sartori G, Arrigoni G, Casonato S, Ruzzene M, Pinna LA, Carignani G.**

**Lyska D, Meierhoff K, Westhoff P.** *How to build functional thylakoid membranes: from plastid transcription to protein complex assembly.* Planta. 2013 Feb;237(2):413-28.

**Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H.** *MODOMICS: a database of RNA modification pathways--2013 update.* Nucleic Acids Res. 2013 Jan;41(Database issue):D262-7

- Magrum LJ, Luehrsen KR, Woese CR.** *Are extreme halophiles actually "bacteria"?* J Mol Evol. 1978 May 12;11(1):1-8.
- Makarova KS, Koonin EV.** *Archaeal ubiquitin-like proteins: functional versatility and putative ancestral involvement in tRNA modification revealed by comparative genomic analysis.* Archaea. 2010 Sep 26;2010.
- Mao DY, Neculai D, Downey M, Orlicky S, Haffani YZ, Ceccarelli DF, Ho JS, Szilard RK, Zhang W, Ho CS, Wan L, Fares C, Rumpel S, Kurinov I, Arrowsmith CH, Durocher D, Sicheri F.** *Atomic structure of the KEOPS complex: an ancient protein kinase-containing molecular machine.* Mol Cell. 2008 Oct 24;32(2):259-75.
- Mao DY, Neculai D, Downey M, Orlicky S, Haffani YZ, Ceccarelli DF, Ho JS, Szilard RK, Zhang W, Ho CS, Wan L, Fares C, Rumpel S, Kurinov I, Arrowsmith CH, Durocher D, Sicheri F.** *Atomic structure of the KEOPS complex: an ancient protein kinase-containing molecular machine.* Mol Cell. 2008 Oct 24;32(2):259-75
- Martin M., Russier M., Brault C., Ouva M., Reynaud JM,** *Mycoplasma contamination enhances stress in cell biology PhD student life. An overview.* PlosLyon,2012, (43:12), 1-16
- Mayr E.** *Two empires or three?* Proc Natl Acad Sci U S A. 1998 Aug 18;95(17):9720-3
- McCloskey JA, Liu XH, Crain PF, Bruenger E, Guymon R, Hashizume T, Stetter KO.** *Posttranscriptional modification of transfer RNA in the submarine hyperthermophile Pyrolobus fumarii.* Nucleic Acids Symp Ser. 2000;(44):267-8.
- Meng FL, Chen XF, Hu Y, Tang HB, Dang W, Zhou JQ.** *Sua5p is required for telomere recombination in Saccharomyces cerevisiae.* Cell Res. 2010 Apr;20(4):495-8.
- Miyauchi K, Kimura S, Suzuki T.** *A cyclic form of N6-threonylcarbamoyladenine as a widely distributed tRNA hypermodification.* Nat Chem Biol. 2013 Feb;9(2):105-11.
- Miyoshi A, Kito K, Aramoto T, Abe Y, Kobayashi N, Ueda N.** *Identification of CGI-121, a novel PRPK (p53-related protein kinase)-binding protein.* Biochem Biophys Res Commun. 2003 Apr 4;303(2):399-405
- Morin A, Auxilien S, Senger B, Tewari R, Grosjean H.** *Structural requirements for enzymatic formation of threonylcarbamoyladenine (t6A) in tRNA: an in vivo study with Xenopus laevis oocytes.* RNA. 1998 Jan;4(1):24-37.
- Morris RC, Elliott MS.** *Queuosine modification of tRNA: a case for convergent evolution.* Mol Genet Metab. 2001 Sep-Oct;74(1-2):147-59.
- Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S.** *Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification.* Nature. 1988 Nov 10;336(6195):179-81.
- Mushegian AR, Koonin EV.** *A minimal gene set for cellular life derived by comparison of complete bacterial genomes.* Proc Natl Acad Sci U S A. 1996 Sep 17;93(19):10268-73.
- Na JG, Pinto I, Hampsey M.** *Isolation and characterization of SUA5, a novel gene required for normal growth in Saccharomyces cerevisiae.* Genetics. 1992 Aug;131(4):791-801.
- Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M.** *The 160-kilobase genome of the bacterial endosymbiont Carsonella.* Science. 2006 Oct 13;314(5797):267.
- Naor A, Thiaville PC, Altman-Price N, Cohen-Or I, Allers T, de Crécy-Lagard V, Gophna U.** *A genetic investigation of the KEOPS complex in halophilic Archaea.* PLoS One. 2012;7(8):e43013.
- Natale DA, Galperin MY, Tatusov RL, Koonin EV.** *Using the COG database to improve gene recognition in complete genomes.* Genetica. 2000;108(1):9-17.
- Nesin M, Lupski JR, Svec P, Godson GN.** *Possible new genes as revealed by molecular analysis of a 5-kb Escherichia coli chromosomal region 5' to the rpsU-dnaG-rpoD macromolecular-synthesis operon.* Gene. 1987;51(2-3):149-61.
- Nichols CE, Johnson C, Lockyer M, Charles IG, Lamb HK, Hawkins AR, Stammers DK.** *Structural characterization of Salmonella typhimurium YeaZ, an M22 O-sialoglycoprotein endopeptidase homolog.* Proteins. 2006 Jul 1;64(1):111-23.
- Nichols CE, Lamb HK, Thompson P, Omari KE, Lockyer M, Charles I, Hawkins AR, Stammers DK.** *Crystal structure of the dimer of two essential Salmonella typhimurium proteins, YgjD & YeaZ and calorimetric evidence for the formation of a ternary YgjD-YeaZ-YjeE complex.* Protein Sci. 2013 Mar 8.
- Nichols CE, Lamb HK, Thompson P, Omari KE, Lockyer M, Charles I, Hawkins AR, Stammers DK.** *Crystal structure of the dimer of two essential Salmonella typhimurium proteins, YgjD & YeaZ and calorimetric evidence for the formation of a ternary YgjD-YeaZ-YjeE complex.* Protein Sci. 2013 Mar 8.
- Nirenberg M.** *Protein synthesis and the RNA code.* Harvey Lect. 1965;59:155-85. Review

**Noma A, Kirino Y, Ikeuchi Y, Suzuki T.** *Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA.* EMBO J. 2006 May 17;25(10):2142-54.

**Oberto J, Breuil N, Hecker A, Farina F, Brochier-Armanet C, Culetto E, Forterre P.** *Qri7/OSGEPL, the mitochondrial version of the universal Kae1/YgjD protein, is essential for mitochondrial genome maintenance.* Nucleic Acids Res. 2009 Sep;37(16):5343-52.

**Oberto J.** *SyntTax: a web server linking synteny to prokaryotic taxonomy.* BMC Bioinformatics. 2013 Jan 16;14:4.

**Onodera T, Satoh K, Ohta T, Narumi I.** *Deinococcus radiodurans YgjD and YeaZ are involved in the repair of DNA cross-links.* Extremophiles. 2013 Jan;17(1):171-9.

**Pace NR, Sapp J, Goldenfeld N.** *Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life.* Proc Natl Acad Sci U S A. 2012 Jan 24;109(4):1011-8.

**Parthasarathy R, Ohrt JM, Chheda GB.** *Conformation and possible role of hypermodified nucleosides adjacent to 3'-end of anticodon in tRNA: N-(purin-6-ylcarbamoyl)-L-threonine riboside.* Biochem Biophys Res Commun. 1974 Sep 9;60(1):211-8.

**Parthier C, Görlich S, Jaenecke F, Breithaupt C, Bräuer U, Fandrich U, Clausnitzer D, Wehmeier UF, Böttcher C, Scheel D, Stubbs MT.** *The O-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction.* Angew Chem Int Ed Engl. 2012 Apr 23;51(17):4046-52.

**Parthier C, Görlich S, Jaenecke F, Breithaupt C, Bräuer U, Fandrich U, Clausnitzer D, Wehmeier UF, Böttcher C, Scheel D, Stubbs MT.** *The O-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction.* Angew Chem Int Ed Engl. 2012 Apr 23;51(17):4046-52

**Peggion C, Lopreiato R, Casanova E, Ruzzene M, Facchin S, Pinna LA, Carignani G, Sartori G.** *Phosphorylation of the Saccharomyces cerevisiae Grx4p glutaredoxin by the Bud32p kinase unveils a novel signaling pathway involving Sch9p, a yeast member of the Akt / PKB subfamily.* FEBS J. 2008 Dec;275(23):5919-33.

**Petkun S, Shi R, Li Y, Asinas A, Munger C, Zhang L, Waclawek M, Soboh B, Sawers RG, Cygler M.** *Structure of hydrogenase maturation protein HypF with reaction intermediates shows two active sites.* Structure. 2011 Dec 7;19(12):1773-83.

**Phillips G, Swairjo MA, Gaston KW, Bailly M, Limbach PA, Iwata-Reuyl D, de Crécy-Lagard V.** *Diversity of archaeosine synthesis in crenarchaeota.* ACS Chem Biol. 2012 Feb 17;7(2):300-5.

**Phizicky EM, Hopper AK.** *tRNA biology charges to the front.* Genes Dev. 2010 Sep 1;24(17):1832-60.

**Powers DM, Peterkofsky A.** *Biosynthesis and specific labeling of N-(purin-6-ylcarbamoyl)threonine of Escherichia coli transfer RNA.* Biochem Biophys Res Commun. 1972 Jan 31;46(2):831-8.

**Powers DM, Peterkofsky A.** *The presence of N-(purin-6-ylcarbamoyl)threonine in transfer ribonucleic acid species whose codons begin with adenine.* J Biol Chem. 1972 Oct 25;247(20):6394-401.

**Prangishvilli D, Zillig W, Gierl A, Biesert L, Holz I.** *DNA-dependent RNA polymerase of thermoacidophilic archaeobacteria.* Eur J Biochem. 1982 Mar 1;122(3):471-7.

**Rajagopala SV, Yamamoto N, Zweifel AE, Nakamichi T, Huang HK, Mendez-Rios JD, Franca-Koh J, Boorgula MP, Fujita K, Suzuki K, Hu JC, Wanner BL, Mori H, Uetz P.** *Escherichia coli K-12 ORFeome: a resource for comparative molecular microbiology.* BMC Genomics. 2010 Aug 11;11:470

**Ranaei-Siadat E, Fabret C, Seijo B, Dardel F, Grosjean H, Nonin-Lecomte S.** *RNA-methyltransferase TrmA is a dual-specific enzyme responsible for C ( 5 ) -methylation of uridine in both tmRNA and tRNA.* RNA Biol. 2013 Apr 1;10(4).

**Reddy DM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Stetter KO, Widdel F, McCloskey JA.** *Structure determination of two new amino acid-containing derivatives of adenosine from tRNA of thermophilic bacteria and archaea.* Nucleic Acids Res. 1992 Nov 11;20(21):5607-15.

**Reddy RP, Schweizer MP, Chheda GB.** *Interaction of the tRNA-derived unusual nucleoside, t6A, with Mn(II) and Mg(II).* FEBS Lett. 1979 Oct 1;106(1):63-6.

**Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W.** *The genome sequence of the thermoacidophilic scavenger Thermoplasma acidophilum.* Nature. 2000 Sep 28;407(6803):508-13.

**Sagan L.** *On the origin of mitosing cells* J Theor Biol. 1967 Mar;14(3):255-74.

**Sankaranarayanan R, Dock-Bregeon AC, Rees B, Bovee M, Caillet J, Romby P, Francklyn CS, Moras D.** *Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase.* Struct Biol. 2000 Jun;7(6):461-5.

**Saraste M, Sibbald PR, Wittinghofer A.** *The P-loop—a common motif in ATP- and GTP-binding proteins.* Trends Biochem Sci. 1990 Nov;15(11):430-4. Review.

**Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S.** *Complete structure of the chloroplast genome of Arabidopsis thaliana.* DNA Res. 1999 Oct 29;6(5):283-90.

**Schweizer MP, Chheda GB, Baczynskyj L, Hall RH.** *Aminoacyl nucleosides. VII. N-(Purin-6-ylcarbamoyl)threonine. A new component of transfer ribonucleic acid.* Biochemistry. 1969 Aug;8(8):3283-9

**Schweizer MP, McGrath K, Baczynskyj L.** *The isolation and characterization of N-[9-(beta-D-ribofuranosyl)-purin-6-ylcarbamoyl]glycine from yeast transfer RNA.* Biochem Biophys Res Commun. 1970 Sep 10;40(5):1046-52.

**Seif E, Hallberg BM.** *RNA-protein mutually induced fit: structure of Escherichia coli isopentenyl-tRNA transferase in complex with tRNA(Phe).* J Biol Chem. 2009 Mar 13;284(11):6600-4.

**Seki Y, Ikeda S, Kiyohara H, Ayabe H, Seki T, Matsui H.** *Sequencing analysis of a putative human O-sialoglycoprotein endopeptidase gene (OSGEP) and analysis of a bidirectional promoter between the OSGEP and APEX genes.* Gene. 2002 Feb 20;285(1-2):101-8.

**Senger B, Auxilien S, Englisch U, Cramer F, Fasiolo F.** *The modified wobble base inosine in yeast tRNA<sup>Ala</sup> is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase.* Biochemistry. 1997 Jul 8;36(27):8269-75.

**Shomura Y, Higuchi Y.** *Structural basis for the reaction mechanism of S-carbamoylation of HypE by HypF in the maturation of [NiFe]-hydrogenases.* J Biol Chem. 2012 Aug 17;287(34):28409-19

**Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schönfisch B, Perschil I, Chacinska A, Guiard B, Rehling P, Pfanner N, Meisinger C.** *The proteome of Saccharomyces cerevisiae mitochondria.* Proc Natl Acad Sci U S A. 2003 Nov 11;100(23):13207-12.

**Smith RF, King KY.** *Identification of a eukaryotic-like protein kinase gene in Archaeobacteria.* Protein Sci. 1995 Jan;4(1):126-9

**Söll D, Ohtsuka E, Jones DS, Lohrmann R, Hayatsu H, Nishimura S, Khorana HG.** *Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids.* Proc Natl Acad Sci U S A. 1965 Nov;54(5):1378-85.

**Sprinzl M, Vassilenko KS.** *Compilation of tRNA sequences and sequences of tRNA genes.* Nucleic Acids Res. 2005 Jan 1;33(Database issue):D139-40.

**Srinivasan M, Mehta P, Yu Y, Prugar E, Koonin EV, Karzai AW, Sternglanz R.** *The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A.* EMBO J. 2011 Mar 2;30(5):873-81.

**Stanier RY, Van Niel CB.** *The concept of a bacterium.* Arch Mikrobiol. 1962;42:17-35.

**Stocchetto S, Marin O, Carignani G, Pinna LA.** *Biochemical evidence that Saccharomyces cerevisiae YGR262c gene, required for normal growth, encodes a novel Ser/Thr-specific protein kinase.* FEBS Lett. 1997 Sep 1;414(1):171-5.

**Stuart JW, Gdaniec Z, Guenther R, Marszalek M, Sochacka E, Malkiewicz A, Agris PF.** *Functional anticodon architecture of human tRNA<sup>Lys3</sup> includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t6A.* Biochemistry. 2000 Nov 7;39(44):13396-404.

**Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD, van Wijk KJ.** *PPDB, the Plant Proteomics Database at Cornell.* Nucleic Acids Res. 2009 Jan;37(Database issue):D969-74

**Suzuki T, Nagao A, Suzuki T.** *Human mitochondrial diseases caused by lack of taurine modification in mitochondrial tRNAs.* Wiley Interdiscip Rev RNA. 2011 May-Jun;2(3):376-86.

**Takemura S, Murakami M, Miyazaki M.** *Nucleotide sequence of isoleucine transfer RNA from Torulopsis utilis.* J Biochem. 1969 Mar;65(3):489-91.

**Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA.** *The COG database: an updated version includes eukaryotes.* BMC Bioinformatics. 2003 Sep 11;4:41.

**Tatusov RL, Koonin EV, Lipman DJ.** *A genomic perspective on protein families.* Science. 1997 Oct 24;278(5338):631-7.

**Teplova M, Tereshko V, Sanishvili R, Joachimiak A, Bushueva T, Anderson WF, Egli M.** *The structure of the yrdC gene product from Escherichia coli reveals a new fold and suggests a role in RNA binding.* Protein Sci. 2000 Dec;9(12):2557-66.

**Teplyakov A, Obmolova G, Tordova M, Thanki N, Bonander N, Eisenstein E, Howard AJ, Gilliland GL.** *Crystal structure of the YjeE protein from Haemophilus influenzae: a putative Atpase involved in cell wall synthesis.* Proteins. 2002 Aug 1;48(2):220-6.

**Thomm M.** *Archaeal transcription factors and their role in transcription initiation.* FEMS Microbiol Rev. 1996 May;18(2-3):159-71.

**Ueda T, Watanabe K, Ohta T.** *Structural analysis of bovine mitochondrial tRNA<sup>Ser</sup> (AGY).* Nucleic Acids Symp Ser. 1983;(12):141-4.

**Valverde R, Edwards L, Regan L.** *Structure and function of KH domains.* FEBS J. 2008 Jun;275(11):2712-26.

**Vendeix FA, Murphy FV 4th, Cantara WA, Leszczyńska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF.** *Human tRNA<sup>Lys3</sup>(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism.* J Mol Biol. 2012 Mar 2;416(4):467-85

**Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M, Kretz K, Lin X, Mathur E, Ni J, Podar M, Richardson T, Sutton GG, Simon M, Soll D, Stetter KO, Short JM, Noordewier M.** *The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism.* Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12984-8.

**Wei FY, Suzuki T, Watanabe S, Kimura S, Kaitsuka T, Fujimura A, Matsui H, Atta M, Michiue H, Fontecave M, Yamagata K, Suzuki T, Tomizawa K.** *Deficit of tRNA<sup>Lys</sup> modification by Cdkal1 causes the development of type 2 diabetes in mice.* J Clin Invest. 2011 Sep;121(9):3598-608

**Whittaker R. H.,** *New concepts of kingdoms of organisms* , Science, vol. 163, 1969, p. 150–160

**Woese CR, Fox GE.** *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proc Natl Acad Sci U S A. 1977 Nov;74(11):5088-90.

**Woese CR, Gupta R, Hahn CM, Zillig W, Tu J.** *The phylogenetic relationships of three sulfur dependent archaebacteria.* Syst Appl Microbiol. 1984;5:97-105.

**Woese CR, Kandler O, Wheelis ML.** *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A. 1990 Jun;87(12):4576-9.

**Woese CR, Magrum LJ, Fox GE.** *Archaeobacteria.* J Mol Evol. 1978 Aug 2;11(3):245-51.

**Woese CR, Maniloff J, Zablen LB.** *Phylogenetic analysis of the mycoplasmas.* Proc Natl Acad Sci U S A. 1980 Jan;77(1):494-8.

**Woese CR, Olsen GJ, Ibba M, Söll D.** *Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process.* Microbiol Mol Biol Rev. 2000 Mar;64(1):202-36.

**Woese CR.** *Bacterial evolution.* Microbiol Rev. 1987 Jun;51(2):221-71

**Woese CR.** *Endosymbionts and mitochondrial origins.* J Mol Evol. 1977 Nov 25;10(2):93-6.

**Woese CR.** *There must be a prokaryote somewhere: microbiology's search for itself.* Microbiol Rev. 1994 Mar;58(1):1-9.

**Wolf YI, Makarova KS, Yutin N, Koonin EV.** *Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer.* Biol Direct. 2012 Dec 14;7:46.

**Xu Q, McMullan D, Jaroszewski L, Krishna SS, Elsliger MA, Yeh AP, Abdubek P, Astakhova T, Axelrod HL, Carlton D, Chiu HJ, Clayton T, Duan L, Feuerhelm J,**

**Yang D, Kusser I, Köpke AK, Koop BF, Matheson AT.** *The structure and evolution of the ribosomal proteins encoded in the spc operon of the archaeon (Crenarchaeota) Sulfolobus acidocaldarius.* Mol Phylogenet Evol. 1999 Jul;12(2):177-85.

**Yarian C, Townsend H, Czystkowski W, Sochacka E, Malkiewicz AJ, Guenther R, Miskiewicz A, Agris PF.** *Accurate translation of the genetic code depends on tRNA modified nucleosides.* J Biol Chem. 2002 May 10;277(19):16391-5.

**Zhang R, Lin Y.** *DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes.* Nucleic Acids Res. 2009 Jan;37(Database issue):D455-8.

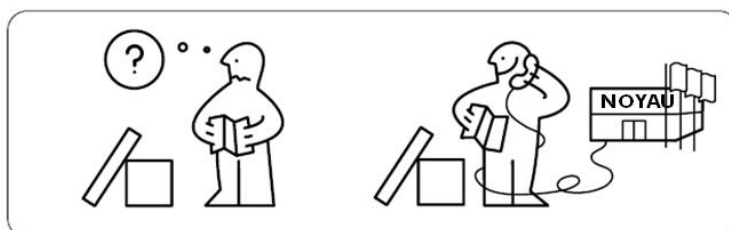
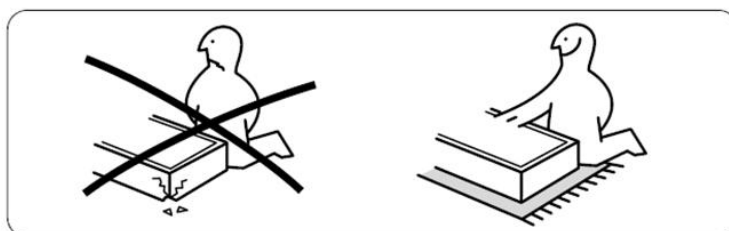
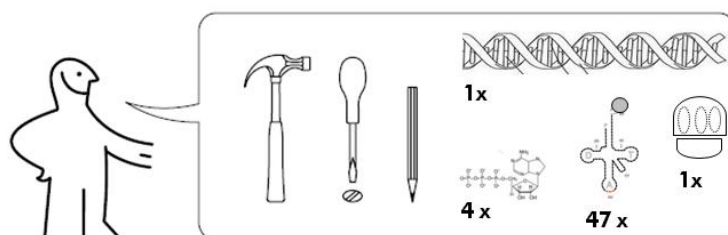
**Zhang R, Ou HY, Zhang CT.** *DEG: a database of essential genes.* Nucleic Acids Res. 2004 Jan 1;32(Database issue):D271-2

**Zhou C, Huang RH.** *Crystallographic snapshots of eukaryotic dimethylallyltransferase acting on tRNA: insight into tRNA recognition and reaction mechanism.* Proc Natl Acad Sci U S A. 2008 Oct 21;105(42):16142-7.

**Zuckerkandl E, Pauling L.** *Molecules as documents of evolutionary history.* J Theor Biol. 1965 Mar;8(2):357-66

# Matériel et Méthodes

## PRÔTEINES



MANUEL DE FABRICATION DE PRÔTEINES





**Annexe 1 : guide très utile pour lire et interpréter les articles scientifiques.**

**Annexe 2 : construction et clonage, et liste des oligonucléotides de synthèse utilisés.**

**Annexe 3 : Séquence et propriétés biochimiques théorique des protéines.**

**Annexe 4 : Tampons et protocoles divers**

### Clonage par synthèse de gène :

**Cgi121-His :** La séquence du gène PAB1522 de *Pyrococcus abyssi* GE5, codant la protéine Cgi121, a été synthétisée chimiquement par GenScript, fusionnée en position 3' à une séquence codant une étiquette hexahistidine, et clonée entre les sites *NdeI*/*NotI* dans le vecteur d'expression pET26(b)+ (Novagen).

**YjeE-StrepTagII :** La séquence du gène codant la protéine YjeE d'*Escherichia coli* (souche K12 MG1665) a été synthétisée chimiquement par GenScript, fusionnée en position 3' à une séquence codant une étiquette StrepTagII, et clonée dans le vecteur d'expression pET26(b)+ entre les sites de restriction *NdeI* et *NotI*.

**Sua5ΔC-ter-StrepTagII et Sua5ΔN-ter-StrepTagII** La séquence correspondant aux résidus 1-209 (domaine N-terminal « YrdC ») et la séquence correspondant aux résidus 230-340 (domaine C-terminal « Sua5 ») de la protéine PAB1303 (Sua5) de *Pyrococcus abyssi* (souche GE5), ont été synthétisées chimiquement par GenScript, fusionnées à une séquence codant une étiquette StrepTagII en position 3' et respectivement clonées dans un plasmide d'expression pET28(a)+, aboutissant aux vecteurs d'expression pET28 :: Sua5ΔCter-StrepTagII et pET28 :: Sua5ΔNter-StrepTagII.

**KEOPS-His (P<sup>H</sup>KBC) :** Une séquence polycistronique comprenant les gènes codant les sous-unités de KEOPS : PAB1159 (Kae1), PAB1047 (Bud32), PAB1522 (Cgi121) et PAB3073 (Pcc1) a été synthétisée chimiquement par Genscript. Une séquence correspondant au site de fixation du ribosome d'*Escherichia coli* (RBS) (5'-TTAACTTTAAGAAGGAG) a été placée en amont de chaque séquence codante. Une

séquence codant une étiquette hexa-histidine a été fusionnée en position 3' du gène PAB3073. La séquence du promoteur T7/LacO et la séquence du terminateur T7 ont été placées respectivement en amont et en aval du polycistron. Cette construction a été clonée entre les sites *Bgl*II et *B*/pI du plasmide pET28(a)+ (Novagen), aboutissant au plasmide d'expression pET28::KEOPS(His).

**PKB-His** : De manière similaire, une construction semblable a été synthétisée en omettant le gène codant la protéine Cgi121.

### Clonages par PCR :

**YgjD-His, YeaZ-His, YrdC-His** : La séquence des gènes codant respectivement les protéines YgjD, YeaZ, et YrdC d'*Escherichia coli* (K12 MG1665) ont été amplifiées par PCR à partir d'ADN génomique d'*Escherichia coli*, en utilisant des couples d'amorces permettant d'ajouter un site de restriction en amont (5') et une séquence codant une étiquette hexahistidine en position 3' suivie d'un site de restriction en aval (3') de chaque gène.

**Pcc1-His, Bud32-His** : La séquence des gènes codants les protéines Pcc1 et Bud32 de *Pyrococcus abyssi* (GE5) ont été amplifiés par PCR à partir respectivement du plasmide pET28 ::KEOPS(His) et de l'ADN génomique de *Pyrococcus abyssi* GE5, selon le même principe.

**BC-His, KB-His, KBC-His** : La séquence des polycistrons respectifs a été amplifié par PCR à partir du plasmide pET28 :KEOPS(His) à l'aide couples d'amorces adaptées (voire annexe), puis clonée entre les sites *Nde*I/*Xho*I dans un plasmide pET26b. La séquence codant l'étiquette hexahistidine en 3' est celle situé sur le plasmide, en aval du site *Xho*I.

**Sua5-His et Sua5-StrepTagII :** La séquence du gène codant la protéine Sua5 (PAB1302) de *Pyrococcus abyssi* (GE5) a été amplifiée par PCR à partir d'ADN génomique de *Pyrococcus abyssi*, en utilisant des couples d'amorces permettant d'ajouter un site de restriction en amont (5') et une séquence codant une étiquette hexahistidine ou une étiquette StrepTagII, suivie d'un site de restriction en aval du gène (3').

**Qri7-His :** La séquence correspondant aux résidus 31 à 407 de la protéine Qri7 de *Saccharomyces cerevisiae* (Y288C), a été amplifiée par PCR à partir d'ADN génomique de *S.cerevisiae*, en utilisant un couple d'amorces permettant d'ajouter un site de restriction en amont (5') et une étiquette hexahistidine suivie d'un site de restriction en aval (3') du gène.

-----

Les produits PCR ont été purifiés, digérés avec des enzymes de restriction adaptées et clonés au niveau du site de clonage multiple (MCS) d'un vecteur de clonage pUC. Le produit de ligation a été transformé dans une souche Top10 d'*Escherichia coli* chimiocompétente, et les clones transformants ont été sélectionnés sur milieu solide LB-Agar additionné d'ampicilline 100µg/mL. Le vecteur recombinant de différentes colonies isolées a été purifié par MiniPrep (kit Wizard Plus Promega), et la séquence présente au sein du MCS a été séquencée (Sigma). Chacun des vecteurs a été digéré avec les enzymes de restriction adaptées, et le produit de digestion correspondant au gène d'intérêt a été purifié (kit NucleoSpin Extract II, Macherey Nagel) puis cloné au niveau du site de clonage multiple (MCS) d'un vecteur d'expression pET. Le produit de ligation a été transformé dans une souche Top10, les clones transformants ont été sélectionnés sur milieu solide LB Agar additionné de kanamycine, et chaque vecteur recombinant a été purifié par MiniPrep.

### Clonage par digestion/ligation (Pcc1, PK)

**Pcc1-His:** Le vecteur pET28a:: KEOPS-His a été digéré par les enzymes *NdeI/XhoI*, permettant d'obtenir un fragment d'ADN correspondant au gène PAB3073 fusionné avec

une séquence codant une étiquette hexahistidine en position 3'. Le produit de digestion a été purifié sur gel et clone dans un vecteur pET26b(+) entre les sites de restriction correspondants.

**PK-His:** Le vecteur pET28a :: KEOPS-His a été digéré par les enzymes *Xba*I et *Sal*I, permettant d'extraire un fragment d'ADN correspondant aux gènes codant Bud32 et Cgi121. Le plasmide ainsi digéré (où se trouvent les gènes codant Kae1 et Pcc1-His) a été purifié sur gel, puis les extrémités cohésives ont été complétées en extrémités franches (blunt) par le fragment de Klenow (NEB), puis ligaturée par la ligase T4 DNA Ligase (NEB).

### Production de protéines recombinantes par sur-expression dans *Escherichia coli* :

Chaque vecteur recombinant a été transformé dans une souche Rosetta2(DE3)pLysS d'*Escherichia coli* (Novagen) rendue électro-compétente ou chimio-compétente. Une colonie isolée a été sélectionnée pour inoculer une pré-culture de 10 à 20mL de LB contenant 50µg/mL de kanamycine et 34µg/mL de chloramphénicol, incubée à 37°C pendant 6 heures à 180rpm. Cette pré-culture a été diluée dans 1 à 2L de milieu OverNight Express Instant TB Express + (Novagen), supplémenté de 10% de glycérol et des antibiotiques précédemment décrits. La culture a été incubée pendant une nuit à 37°C, 180rpm, permettant la croissance des cellules parallèlement à la sur-expression des protéines recombinantes.

Dans le cas des protéines Kae1, Cgi121, Bud32, Pcc1, YgjD, YeaZ, YrdC, un protocole alternatif d'induction a été utilisé. Une colonie isolée a servi à inoculer une pré-culture de 100mL de LB (supplémentée des antibiotiques de sélection), incubée pendant une nuit à 37°C, 180rpm. Cette préculture a été diluée dans 1 à 2L de LB + antibiotiques, à une densité optique (D.O.) finale de 0.1, et incubée à 37°C, 180rpm. A DO = 0,7, la sur-expression des protéines recombinantes a été induite par addition de 1mM d'IPTG, pendant 4 heures à 37°C, 180rpm.

Les cellules ont été collectées par centrifugation à 5.000xg pendant 10mn à 4°C, et le culot cellulaire a été congelé et stocké à -20°C jusqu'à utilisation.

**Note 1** : Dans les cas des protéines YgjD, YeaZ, YrdC, YjeE, la souche d'expression utilisée est BL21 Escherichia coli (Novagen)

### Purification des protéines recombinantes portant une étiquette hexahistidine par chromatographie d'affinité sur résine Ni-NTA :

Le culot cellulaire a été resuspendu dans un tampon de lyse (Tris-HCl 50mM, pH 8.0, NaCl 200mM/500mM,  $\beta$ -mercaptoéthanol 5mM, glycérol 10%,) supplémenté d'un cocktail d'inhibiteurs de protéase (Roche) pour limiter la dégradation des protéines, de lysozyme à une concentration de 1 mg/mL pour faciliter la lyse, et 10mM d'imidazole pour limiter les interactions non spécifiques lors de la purification. Le volume de tampon (V) additionné est calculé selon la formule de normalisation:  $[V_{\text{Tampon}} = D.O. \times V_{\text{culture}} \times 50 \times 10^{-3}]$ . Les cellules ont été ensuite lysées par sonication sur glace, et l'extrait cellulaire a été centrifugé à 30.000xg pendant 30mn à 4°C pour éliminer les débris cellulaires et les agrégats protéiques insolubles.

Un volume de 2mL de résine nickel nitrotriacetate agarose (Ni-NTA) (Quiagen) a été placée dans une colonne de purification par gravité, puis équilibrée avec 10 CV (volume colonne) de tampon de lyse. L'extrait soluble précédemment obtenu a été purifié par passage sur cette résine, à 4°C. La résine a été lavée par un volume de tampon de lyse équivalent à la moitié du volume utilisé pour la lyse, supplémenté de 20mM d'imidazole.

Les protéines recombinantes ont été éluées par addition de tampon de lyse supplémenté de 100, 200 et 400mM d'imidazole (respectivement par fractions de 4x1mL, 4x1mL et 10x1mL). Les fractions contenant la protéine d'intérêt ont été réunies dans un concentrateur Amicon (Millipore), concentrées par centrifugation à 4.000xg, jusqu'à un volume de 2mL. La taille des pores utilisée (cut-off) est calculée selon la formule :  $[Masse\ moléculaire\ théorique\ de\ la\ protéine\ d'intérêts / 2]$ .

La solution protéique concentrée a été centrifugé pendant 10mn à 20.000xg pour éliminer les précipitats éventuels, et le surnageant à été purifié par chromatographie en filtration sur gel.

**Note 1** : Dans le cas des protéines de *Pyrococcus abyssi*, le surnageant a été chauffé à 65°C pendant 15mn afin de précipiter les protéines solubles de la souche d'expression, puis centrifugé à 30.000xg pendant 30mn, pour éliminer les protéines précipitées.

**Note 2** : Dans le cas des complexe  $P^H$ KBC et de la protéine Qri7, le tampon de lyse contient 500mM de NaCl.

### Purification des protéines recombinantes portant une étiquette StrepTag II (IBA).

Les cellules ont été lysées selon le même protocole dans un tampon équivalent. Le surnageant obtenu a été purifié par passage sur une résine StrepTactin (IBA). Après lavage avec du tampon de lyse, les protéines ont été éluées par addition de tampon de lyse supplémenté de desthiobiotine 2,5 mM, compétiteur de l'étiquette StrepTagII, selon le protocole fourni par le fabricant. La suite de la purification est alors identique à celle décrite précédemment.

### Purification des protéines recombinantes par gel filtration :

La solution protéique concentrée à l'issue de la chromatographie d'affinité à été purifiée par une seconde étape de chromatographie sur gel filtration en FPLC (AKTA) pour éliminer les contaminants résiduels, l'imidazole ou la desthiobiotine. La solution protéique a été injectée sur une colonne SuperDex<sup>TM</sup>200 ou SuperDex<sup>TM</sup>75 16/60 selon la taille théorique de la protéine d'intérêt, et les protéines ont été séparées dans un tampon de purification (TrisHCl 50mM, NaCl 200/500, glycérol 10%) à 4°C. La quantité de protéine dans chaque fraction récoltée a été suivie par mesure de l'absorbance à 280nm.

Les fractions contenant la protéine d'intérêt ont été réunies et concentrées. Après mesure du spectre d'absorbance et détermination de la concentration, la solution protéique a été alicotée, congelée dans de l'azote liquide et stockée à -80°C jusqu'à utilisation.

Les différentes étapes de la purification en chromatographie d'affinité et gel filtration ont été analysées par SDS-PAGE.

### Analyse du spectre d'absorption et détermination de la concentration de protéine

Le spectre d'absorption des protéines purifiées a été mesuré de 200nm à 1000nm à l'aide d'un spectrophotomètre UV-visible Spectrophotometer Cary-50 Scan (Varian). Le coefficient d'extinction théorique de chaque protéine a été calculé par le logiciel ProtParam (Expasy, UniProt), et la concentration protéique a été déduite d'après la valeur de l'absorbance à 280nm.

### Détermination de la concentration protéique par quantification au réactif de Bradford.

La concentration de protéine a également été déterminée par la technique colorimétrie au réactif de Bradford, en utilisant une solution Bio-Rad Protein Assay (Bio-Rad) et une courbe standard obtenue par une gamme de BSA (bovine serum albumin).

### Analyse SDS PAGE des protéines :

Un alicot de chaque extrait protéique a été additionné d'un tampon de dénaturation Laemmli 5X (annexe 4)

Les protéines ont été dénaturées pendant 10mn à 95°C, puis ont été séparées par électrophorèse sur gel d'acrylamide (composition : annexe 4) dans un tampon de migration contenant Tris Glycine SDS (annexe 4), sous une intensité de courant constante de 30mA. Un mélange de protéines colorées (Sigma, n°671) a été utilisé

comme marqueur de poids moléculaire.. Après migration, le gel a été rincé dans de l'eau MilliQ, puis coloré avec une solution Instant Blue™ (Expedeon) basé sur le principe de la coloration au bleu de Coomassie, pendant 1h à 1 nuit. La coloration aspécifique du gel a été éliminée par rinçages répétés dans de l'eau MilliQ une nuit minimum.

### Analyse des acides nucléiques liés aux protéines en gel d'agarose :

Un alicot de chaque extrait protéique a été additionné d'un tampon de dénaturation Laemmli et dénaturé comme dans le cas d'une analyse SDS PAGE. La solution dénaturée a été séparée par électrophorèse sur un gel d'agarose 1 à 2%, contenant 0,5µg/mL de Bromure d'éthidium, dans un tampon TAE 0.5X (annexe 4). La présence d'acide nucléique dans le gel a été révélée par exposition aux UV.

### Construction des plasmides de sur-expression des ARNt :

La séquence des gènes d'ARNt d'*Escherichia coli* (Ec) et de *Pyrococcus abyssi* (Pa) : PaVal (anticodon : GUC), PaLys (AAA), PaLeu (AUC), PaAsn (AAC), PaMet (ATG), EcVal (GUC), EcLys (AAA), EcIle (AUC), EcAsn (AAC) ont été extraits de la banque de données Genomic tRNA Database (<http://gtrnadb.ucsc.edu/>) (Lowe and Eddy, 1997).

Deux oligonucléotides déphosphorylés au niveau de l'extrémité 5', et correspondant aux séquences sens et anti-sens des gènes d'ARNt ont été synthétisés par la société Integrated DNA Technologies.

Un mélange d'oligonucléotides sens et anti-sens (50µM) a été incubé à 90°C et refroidi lentement, afin de permettre l'appariement des deux fragments d'ADN. Le fragment d'ADN double brin portant des extrémités cohésives AATT et TGCA en position 5' et 3' a été cloné dans un vecteur pBlue-Script digéré par les enzymes *EcoRI* et *PstI*, aboutissant aux vecteurs pBS-ARNt.

Les gènes d'ARNt ainsi clonés sont placés sous contrôle du promoteur T5/opérateur Lac et le terminateur *rrnC*, et leur séquence a été vérifiée par séquençage (Sigma).



## Production d'ARNt par sur-expression dans *Escherichia coli*

Une souche XL1 Blue d'*Escherichia coli* (Agilent Technologies) a été transformée avec les vecteurs d'expression pBS-ARNt décrits précédemment, et sélectionné sur milieu solide LB Agar supplémenté avec de l'ampicilline (100µg/µL). Une colonie isolée a été utilisée pour inoculer une culture de 50mL de milieu OverNight Expressin Instant TB (Novagen), supplémentée avec 10% de glycérol et 100µg/µL d'ampicilline, et incubée durant 16 heures (une nuit + une matinée) à 37°C, 180rpm. Les cellules ont été centrifugées à 5000g pendant 10mn à 4°C. Les ARN totaux ont été extraits avec du TriReagent (Sigma) selon le protocole du fournisseur. La phase aqueuse a été précipité à l'isopropanol, incubé 30min à -20°C. Les ARN totaux précipités ont été centrifugés 30mn à 20.000xg, 4°C, puis le culot a été lavé 3 fois avec de l'éthanol 70%. Après séchage à température ambiante, le culot a été resuspendu dans de l'eau RNase free (Sigma), la concentration en ARN de la solution a été calculée d'après l'absorbance à 260nm (NanoVue), et stockée à -80°C jusqu'à purification.

## Purification des ARNt par la méthode "crush and soak":

La solution d'ARN totaux a été additionnée avec un volume équivalent de tampon de dénaturation contenant 95% formamide (annexe 4), et dénaturée par chauffage à 70°C pendant 10mn. Les ARN ont été séparés par électrophorèse en condition dénaturante sur un gel de 20cm (épaisseur 1mm) contenant 8M urée, TBE1X (annexe 4), acrylamide/bisacrylamide 19 :1 6 à 8% sous une tension constante de 100V dans un tampon TBE1X. Une fois que le xylène cyanol est arrivé à quelques centimètres du bas du gel, celui-ci a été démoulé et la zone contenant les ARNt d'intérêt a été découpée à l'aide de la technique de visualisation (UV shadow ou ombre à l'ultraviolet) suivante :

Le gel, placé sur deux couches de parafilm servant d'écran visuel, a été exposé brièvement par le haut par une lampe UV (courte longueur d'onde 238nm) dans une pièce noire, permettant de visualiser l'ombre créée par la présence des ARN dans le gel. Des encoches ont été rapidement effectuées dans le gel afin de délimiter la zone contenant l'ARNt sur-exprimé, visible sous la forme d'une ombre majoritaire, puis la bande de gel correspondante a été découpée sous lumière visible.

Cette bande de gel a été finement broyé (**crush**) à l'aide d'une pipette pasteur, puis les ARNt ont été élués (**soak**) dans un volume de 10 à 15mL de tampon d'éluion contenant 0,5 M ammonium acetate, 10mM magnesium acetate, 0,1% SDS, 1mM EDTA, pendant une nuit à 37°C avec agitation constante.

Les fragments de polyacrylamide ont été centrifugés 10min à 5000g, lavés deux fois avec du tampon d'éluion, puis les surnageant ont été réunis et filtrés à travers un filtre 0,22µm PES membrane (Millipore). Les ARNt ont été précipités par addition de 2 volumes d'éthanol absolu et incubation à -20°C une nuit. Après centrifugation à 15000g pendant 30min, le culot a été lavé par de l'éthanol 70%, séché et resuspendu dans de l'eau RNase free (Invitrogen).

La concentration d'ARNt a été déterminée par mesure de l'absorbance à 260nm (NanoVue GE Healthcare) et la pureté de l'ARNt a été vérifiée par électrophorèse sur gel d'acrylamide.

**Note** : les ARNt transcrits *in vitro* ont été purifiés selon la même méthode.

## Mesure de l'incorporation *in vitro* de la C14 L-thréonine

**Note** : Dans toutes les expériences, le milieu réactionnel de base indiqué par la suite est identique, mais les autres composants additionnels peuvent être omis ou différents, de même que la température ou le temps d'incubation. Ces informations sont données dans la partie résultat. L'exemple ci-dessous correspond à une expérience type réalisée avec le complexe KEOPS et la protéine Sua5 de *P.abysyi*. Dans la partie II et III des Résultats, le milieu réactionnel de base contient systématiquement 100µM de MnCl<sub>2</sub>.

Un milieu réactionnel de base contenant 50mM de tampon HEPES, pH8.0, 35mM de KCl, 5mM de MgCl<sub>2</sub>, 5mM de DTT, 18,2µM de C14 L-thréonine (0,05 µCi, 55 Ci/mol, Hartmann Analytic), a été additionné de 100µM de MnCl<sub>2</sub>, 1mM d'ATP, 5mM de NaCO<sub>3</sub>, 1µM de KEOPS, 1µM de Sua5, 5 µM d'ARNt dans un volume final de 50µL, et incubé à 55°C pendant 30min. Une goutte de 20µL d'huile minérale (Sigma Aldrich) a été déposée au niveau de la surface du milieu réactionnel afin d'empêcher l'évaporation.

Après incubation, la réaction est stoppée par l'addition de 1mL de TCA 15%, puis incubée sur glace pendant 30min à 1 heure afin de précipiter les macromolécules. Le matériel précipité a été filtré sur un filtre GF/F glass microfiber (Whatmann) pré-humidifié avec 1mL de TCA 15% placé dans un dispositif de filtration sous vide (Millipore). Le tube contenant le mélange réactionnel a été rincé 2 fois avec 1mL de TCA15%, puis le filtre a été lavé par 3 x 1mL de TCA 5% suivi de 3 x 1mL d'éthanol 95%. Le filtre a été alors séché de manière extensive sous une lampe chauffante et hotte aspirante pendant 15min, puis placé dans une fiole contenant 3mL d'un cocktail de scintillation Emulsifier Safe (Zinsser analytic). La radioactivité présente sur le filtre a été mesurée par un détecteur Packard Liquid Scintillation Analyzer comme la moyenne des Coups par minute (CPM) sur une durée de 2min.

La quantité de C14 L-thréonine présente sur le filtre a été calculée à partir d'une courbe de calibration obtenue à l'aide de différentes quantités connues de C14 L-thréonine déposé sur filtre (1 pmol thréonine = CPMx102,840). Le rendement de la réaction (en %) est calculé selon la formule  $[\text{CPM mesuré} - \text{CPM (bruit de fond)}] / 102,840 / (\text{pmol d'ARNt}) \times 100$ .

### Extraction des ARNt après réaction d'incorporation *in vitro*:

Un mélange réactionnel identique au précédent, en l'absence d'huile minérale, a été incubé dans les mêmes conditions, puis la réaction a été stoppée par addition de 450µL d'acétate d'ammonium 0,3M, pH 5.25. Les ARNt ont été extraits par addition d'un mélange iso-volumique d'aquaphénol et de chloroforme/alcool isoamylique (Ready Red). Le mélange a été vortexé 1min, puis centrifugé 15min à vitesse maximum 15.000xg pendant 10min. La phase aqueuse supérieure a été récupérée, et soumise à une nouvelle étape d'extraction. La phase aqueuse obtenue a été additionnée de 250µL de chloroforme pour éliminer toute trace de phénol, centrifugée et ré-isolée.

Les ARNt présents dans la phase aqueuse ont été précipités par addition de 2 volumes d'éthanol absolu et incubation à -20°C pendant 1 heure à 1 nuit. Après centrifugation à 20.000xg pendant 30min, le surnageant a été éliminé, et le culot d'ARNt a été lavé à l'éthanol 70°, puis séché sous hotte pendant 10mn.

## Digestion des ARNt en mononucléotides et séparation par chromatographie sur couche mince en 2D dimensions

Le culot d'ARNt a été resuspendu dans 15µL d'acétate d'ammonium 50mM, additionné de 1 µg de nucléase P<sub>1</sub> de *Penicillium citrinum* (Sigma), permettant de digérer l'ARN en mononucléotides 5'pN. La réaction de digestion a été incubée une nuit dans une étuve à 37°C.

Une solution de 5'p mononucléotides pA, pU, pG, pC froids (concentration 4µg/µL) servant de marqueurs de migration a été obtenue par digestion d'ARNt totaux de *S.cerevisiae*(Sigma) non radiomarqués dans les mêmes conditions.

2µL de la réaction de digestion ont été mélangés avec 2µL à 4µL de la solution de marqueurs (équivalent à 8 à 16µg) et déposé sur une plaque de cellulose 10x10 cm CEL-300 (Merck), par dépôt successifs de 0.5µL. Deux plaques identiques ont été utilisées pour chaque expérience.

Les 5'P-monomononucleotides ont été séparés en utilisant trois solvants de chromatographie différents : solvant A = acide isobutyrique/ammoniac 25%/eau (66/1/33 [v/v/v]), solvant B tampon phosphate/sulfate d'ammonium/propan-1-ol (100/60/2 [v/w/v]) and solvant C isopropanol/HCl 35%/eau (68/18/14 [v/v/v]).

Une première migration verticale a été effectuée dans le solvant A. Une fois que le front de migration a atteint le haut des plaques, celles-ci ont été séchées de manière extensive sous hotte pendant une nuit. La seconde dimension de chromatographie a été réalisée à perpendiculairement à la première migration, pour l'une des plaques dans le solvant B, et pour l'autre plaque dans le solvant C. Après migration, les plaques ont été séchées, et un alicot (0,5µL) de la réaction de digestion initiale a été déposée aux 4 coins de chaque plaque, afin de servir de repère pour aligner les images obtenues après révélation.

La position des 4 nucléotides majoritaires pA, pU, pC, pG a été révélée par exposition aux UV, et la position du nucléotide pt<sup>6</sup>A radiomarké a été révélé par exposition de la plaque de chromatographie sur écran au phosphore (GE), suivi d'une lecture à 533nm par un scanner Typhoon. Les deux images obtenues ont été superposées afin de comparer la position du nucléotide radiomarké par rapport aux cartes de référence.

## Activité ATPase en chromatographie sur couche mince

Un milieu réactionnel de base contenant 50mM de tampon HEPES, pH8.0, 35mM de KCl, 5mM de  $MgCl_2$ , 5mM de DTT (identique au milieu réactionnel utilisé pour la réaction de biosynthèse de  $t^6A$  *in vitro*), a été additionné des protéines d'intérêts (1 $\mu$ M par défaut), supplémenté de 100 $\mu$ M ou 10 $\mu$ M d'ATP froid et 1 $\mu$ Ci d'ATP radioactif [ $\alpha$ P32] ou [ $\gamma$ P32] (par défaut l'ATP<sup>[ $\alpha$ P32]</sup> est utilisé) (Perkin Elmers), dans un volume final de 20 $\mu$ L. D'autres composants réactionnels à différentes concentrations (ARNt, L-thréonine, carbonate,  $MnCl_2$ ) ont été ajoutés selon l'expérience (indiqués dans le texte).

Ce mélange réactionnel a été incubé à différentes températures (par défaut 55°C), durant un temps variable (par défaut, 30mn), puis la réaction a été stoppée sur glace.

Avant dépôt, une première migration dans de l'eau distillée a été préalablement effectuée sur une plaque 10X10cm PEI-Cellulose (ou 20x20 selon les cas) (Merck) afin d'éliminer les éventuelles impuretés. Après séchage de la plaque, 1 $\mu$ L du mélange réactionnel précédent a été déposé à 2cm de la partie inférieure de la plaque, puis le dépôt a été séché à température ambiante. Les nucléotides radioactifs ont été séparés par chromatographie sur couche mince (CCM) en plongeant la plaque verticalement dans un solvant D : 0,5 M  $KH_2PO_3$ - pH 3.5 (ou selon le cas dans un solvant E : LiCl 0,5M, acide formique 1M), sur une hauteur de 1cm. Lorsque le solvant a atteint la partie supérieure de la plaque, celle-ci a été séchée de manière extensive, puis la radioactivité a été révélée par exposition sur un écran PhosphorScreen (GE) et une lecture à 533nm par un scanner Typhoon.

Le solvant D permet de séparer les composés selon l'ordre : dépôt/ATP/ADP/AMP/Pi, le solvant E permet de séparer les composés selon l'ordre dépôt/ATP/PPi/ADP/Pi/AMP.

**Note:** Dans la partie II des résultats, chaque sous complexe est à une concentration de 1 $\mu$ M, l'ATP froid est à une concentration de 10 $\mu$ M, l'ARNtLys(UUU) à une concentration de 2 $\mu$ M, le solvant D est utilisé. Dans les autres cas, les conditions sont mentionnées.

### Test de phosphorylation en SDS PAGE :

Un mélange réactionnel similaire à celui utilisé pour le suivi de l'activité ATPase , contenant 25µM d'ATP froid supplémenté de 1µCi d'ATP [<sup>32</sup>P], a été additionné des protéines et des différents composants indiqués dans le texte, dans un volume finale de 20µL, et incubé à 55°C pendant 30mn. 1µL de la réaction a été analysée sur chromatographie sur couche mince , et le reste de la réaction a été additionnée de 5µL de tampon de dénaturation Laemmli 5X, et dénaturé par chauffage à 95°C pendant 10mn. Les protéines ont été séparées par SDS-PAGE 15%. A l'issue de la migration, les nucléotides fixés de manière aspécifique dans le gel ont été éliminés par un rinçage dans un bain de TCA 10% chauffé à 90°C, suivi d'un rinçage en TBE1X. Les protéines ont été révélées par coloration avec une solution InstantBlue. La radioactivité présente dans le gel a été révélée par *phosphorimaging*.

### Test d'interaction par gel filtration :

Un mélange de 100µg de chaque protéine (combinaisons indiquées dans le texte) a été incubés dans un volume final de 200µL d'un milieu réactionnel contenant un tampon Tris-HCl pH8, NaCl 200mM, glycérol 10% (ou d'autres composants si indication). Après incubation 20min à 4°C, température ambiante ou autre selon indication, la totalité du volume réactionnel a été injecté sur une colonne SuperDex 200 10/300, et les complexes protéiques ont été séparés par chromatographie en gel filtration par FPLC, dans un tampon Tris-HCl pH8, NaCl 200mM, glycérol 10% à 4°C. La quantité de protéine présente dans chaque fraction (0,5mL) a été suivie en sortie de colonne par mesure de l'absorbance à 280nm, et la composition des fractions d'intérêt a été analysée par SDS-PAGE.

### Retard sur gel d'acrylamide en condition native :

#### **Marquage des sondes:**

Les ARNt (ou ARN/ADN simple brin synthétiques) ont été déphosphorylés en incubant 20pmol d'ARN avec 1 unité de Calf Intestinal alkaline phosphatase (CIAP) (Fermentas), dans

un volume de 10µL pendant 1 heure à 37°C, puis extraits par addition de 140µL d'acétate de sodium pH 5.25, 0,3M, et de 150µL d'un mélange d'aquaphénol/chloroforme/alcool isoamylique en proportions 25 :24 :1 (Ready Red).

Le mélange a été vortexé et centrifugé 10min à 20.000xg, et les ARNt présents dans la phase aqueuse ont été précipités par ajout de 2 volumes d'éthanol absolu, et incubé 30mn à -80°C.

Après centrifugation pendant 30mn à 20.000xg, 4°C, le culot d'ARN précipité a été séché à température ambiante, puis resuspendu dans un volume final de 20µL contenant 1µL de T4 Polynucléotide kinase (T4PNK, NEB), un tampon adapté à l'activité phopshotransférase (tampon A, NEB), et 10µCi d'ATP<sup>[γP<sup>32</sup>]</sup> (6000 Ci/mmol) (Perkins), permettant de marquer radioactivement les oligonucléotides en position 5'.

Après incubation à 37°C pendant 1 heure, le volume réactionnel a été ajusté à 50µL avec de l'eau *RNAse free* et l'ATP radioactif non incorporé a été éliminé par gel filtration sur colonne G25 (GE Healthcare) selon le protocole du fournisseur.

La solution d'ARN radioactif obtenue a été chauffé à 75°C pendant 10mn pour dénaturer la T4PNK et assurer un repliement des ARNt, refroidie lentement à température ambiante et stockée à -20°C jusqu'à utilisation.

Dans le cas des sondes 65mer ARN/ADN synthétiques, l'obtention d'une sonde radioactive double brin a été réalisée en hybridant une quantité équimolaire d'oligonucléotide simple brin radiomarké avec un oligonucléotide complémentaire, incubé 10 min. à 70°C puis refroidis lentement jusqu'à température ambiante.

### **Retard sur gel:**

La réaction de fixation a été réalisée dans un volume final de 20µL contenant 0,05pmol (10nM) d'oligonucléotide (ARNt) radiomarkée et de protéine à différentes concentrations, dans un tampon contenant 50mM TrisHCl, pH8, 100mM NaCl, 1mM DTT, 5mM MgCl<sub>2</sub>, glycérol 5%. Après incubation pendant 15min à 4°C ou à température ambiante, 5µL de tampon de dépôt 5X (10 mM Tris-HCl pH 7.5, 20% glycerol, 1 mM EDTA, 0.1 mg/ml BSA, et 0.1% xylène cyanol) ont été ajoutés à la réaction, puis le mélange a été déposé sur un gel de

polyacrylamide (19 :1) 6%, TGE 1X (50 mM Tris, pH 7.5, 8 mM glycine, 0.1 mM EDTA), glycérol 5%, et les complexes nucléo-protéiques ont été séparés par électrophorèse dans un tampon TGE1X, sous une tension de 100V (7.5 V/cm) à température ambiante pendant 3 à 4 heures.

Après migration, le gel a été séché puis la radioactivité a été révélée par exposition sur un écran PhosphorScreen (GE) et une lecture à 533nm par un scanner Typhoon.



### Note :

Ce projet a été mis en place et mené grâce à l'intervention de différents participants, et je tiens à mentionner la contribution respective de chacun:

**A.Hecker** : conception du polycistron KEOPS, clonage des gènes YgjD, YeaZ, YrdC, Qri7, Kae1, Cgi121, Bud32, Pcc1, Sua5. Mise au point de la purification de KEOPS, YgjD, YeaZ, Kae1, Qri7.

**T.Basta** : clonage des ARNt

**E.Crozat** : clonage des ARNt, mise au point de la réaction t<sup>6</sup>A in vitro

**D. Guetta** : clonage, purification et caractérisation des sous complexes de KEOPS

**W.Zhang** : purification des protéines de *S.cerevisiae*

**B.Collinet** : purification des protéines de *S.cerevisiae*

**J. Bareille** : transcription *in vitro* des ARNt



# **ANNEXES**



# Annexe 1: A Guide to Understanding Journal Phraseology

**The author writes:** It has long been known.

**REALLY means:** I haven't bothered to look it up.

**The author writes:** It is believed.

**REALLY means:** I think.

**The author writes:** It is generally believed.

**REALLY means:** A couple of other guys think so too.

**The author writes:** It is not unreasonable to assume.

**REALLY means:** If you believe this, you'll believe anything.

**The author writes:** Of great theoretical importance.

**REALLY means:** I find it kind of interesting.

**The author writes:** Of great practical importance.

**REALLY means:** I can get some mileage out of it.

**The author writes:** Typical results are shown.

**REALLY means:** The best results are shown.

**The author writes:** Three samples were chosen for further study.

**REALLY means:** The others didn't make sense, so we ignored them.

**The author writes:** The 4 hour sample was not studied.

**REALLY means:** I dropped it on the floor.

**The author writes:** The 4 hour determination may not be significant.

**REALLY means:** I dropped it on the floor, but scooped most of it up.

**The author writes:** The significance of these results is unclear.

**REALLY means:** Look at the pretty artifact.

**The author writes:** It has not been possible to provide definitive answers.

**REALLY means:** The experiment was negative, but at least I can publish the data somewhere.

**The author writes:** Correct within an order of magnitude.

**REALLY means:** Wrong

**The author writes:** It might be argued that.

**REALLY means:** I have such a good answer for that objection that I shall now raise it.

**The author writes:** Much additional work will be required.

**REALLY means:** This paper is not very good, but neither are all the others in this miserable field.

**The author writes:** These investigations proved highly rewarding.

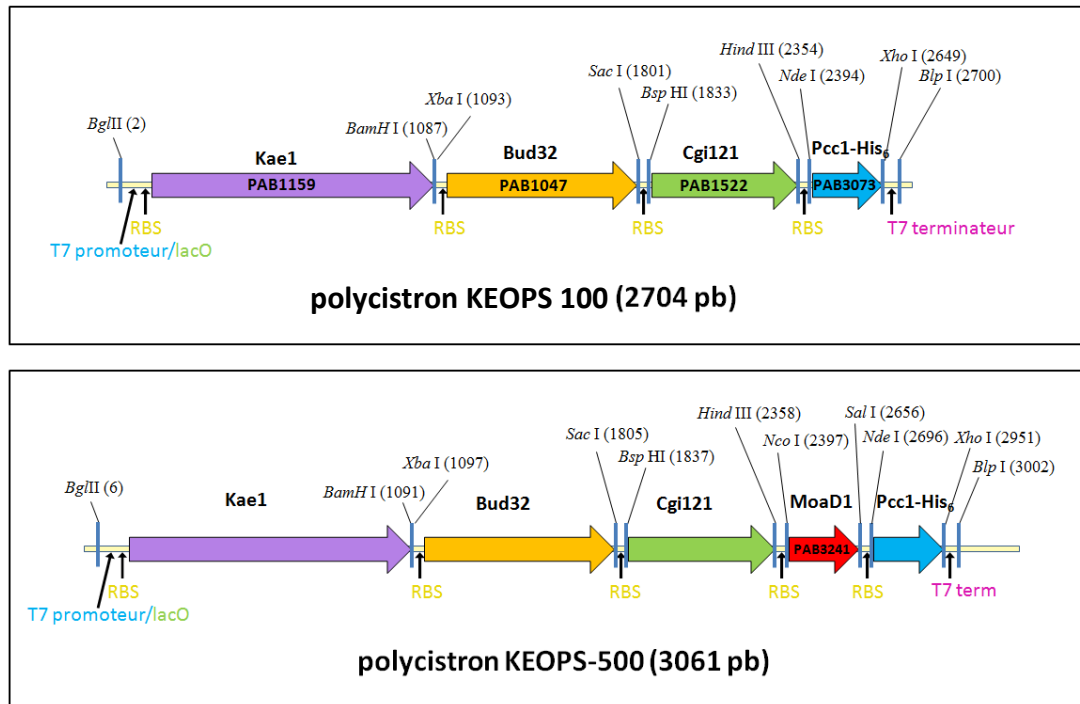
**REALLY means:** My grant is going to be renewed.

**The author writes:** I thank Smith for assistance with the experiments and Jones for useful discussions on the interpretation of the data.

**REALLY means:** Smith did the experiment and Jones explained it to me.

# Annexe 2 : Constructions et clonage

## Carte et séquence des cassettes *Pa*-KEOPS 100 et *Pa*-KEOPS 500



AGATCTCGATCCCGGAAATTAATACGACTCACTATAGGGAATTGTGAGCGGATAACAATTCCCTCTGTAATAATTTTGTTAACCTTAAAGAGGACATATAGTTATGCTAGCCCTGGGCAT  
TGAGGGGACGGCACACACATTGGGAATAGGAATAGTTTCGGAAGATAAAGTATTAGCAATGTTTTTGACACATTAACTGAAAGGGAGGAATCCATCCCAAGAGAGCTCGGAGCATCATG  
CTAGGTTAATGAAGCCCTTGTGAGGAAAGCATTATCTGAGGCTGGAGTTTCGTTGGATGATATAGACGTTATAGCGTTCTCCAGGGACAGGATTGGGACCCGACTTAGGGTTGTCGCTACG  
GCGCGAGGGCTTTGGCGGTTAAATATAGGAAACCGATAGTTGGGGTAAACCACTGTATAGCCATGTAGAGATAACAAGATGTTCCGAGTGAAGACCCAGTTGGCCCTTATGTTAGTGGCGG  
GAACACCCAAAGTTCTGGCCCTTGAAGGAGGAAGGTATAGGGTATTGGAGAGACCCATAGATATAGGAATCGGAACGCCATAGATGCTTTGCAAGGGAACCTCGGATTGGGGTTCCCTGGGGGGC  
CAAGGTAGAGAACTAGCAGAGAAAGGGGAGAAATACATTGAGTTACCATACGCCGTTAAAGGCATGGACTTATCCTTTCTGGACTCTTAACGGAAGCCATTAGGAAGTATAGAAAGTGAAGAA  
TACAGGGTAGAGGATTTAGCGTATTCATTTTCAAGAACTGCATTTGCGGCATTAGTTGAGGTAAACGAAAGGGCAGTTGCCACACGGGAAAGATGAGGTAGTCTCGTTGGAGGGGTTGCAGC  
GAACAATAGGCTTAGGGAATGCTCAGAATAATGACAGAGGATAGAGGTATAAAGTTCTTCGCTCCCTCTTATGACTTGTCCCGGATAATGGGGCCATGATAGCTTATACAGGCTTAAGGATGT  
ATAAGCGGGGAATCAGCTTCAGGCTTGAAGAACTATAGTCAAGCAGAAGTTTCAAGGACTGATGAGGTGGAGATTGTATGGTAAGTGGTCTCTAGAAATTTTGTAACTTTTAAAGGAGATA  
TAGTTATGAAGCTAATCAAGCAAGAGCGGAGGCCAAGATTATCTCTGAGAGTTCTCGGAGCTTACTTTGATTATCCAATTAAGGTTATAGTGAAGGAGAGGATTAAAGAGAGGTACAGGATT  
CCAGAGATCGATCTGAAGCTCAGGAAAGAGAGAACTATCAGGAGGCCAAGGATCTCGAAGGGCCAAAGGAATTTGGAGTCAATGTTCCCTACGTGTTTGAAGTTGACACTAAAAATATGATAAT  
AGTGTGAGGTACATAGAGGGGAAAGGCTCAAGAGCTTCTGGAAGAGCTCCCAATGGAAGAAAGACTTAAAGTATGCAGGGAAGTCGGAAGGCAGATTGGAAGTTACATGAGGCGGGATAG  
TTCACGGGACTTAACGACCTCAATATGATACTTAGGGAAGGCAAGGTTTACTTCATAGATTTCGGGCTTGGCGAGTTTGACGATACGATTGAAGCTCAGGAGTGTAGATTACACCTGCTTAAG  
AGAGCAATGGAGAGTACTCATTATAAGTGGTTTGAAGGGGCTTTGAGGAAGTTTGAAGGGATACATAGAAATCAGAGGAGAAGATAAGGGAAGGAAATTAGGGAGAAGATAAGGGAAATAGA  
ATTGAGGGGTAGGTACAGGGAGAGAGCTGGATTACGCGAGTAAGTCTGAGCTCAATTTTGTTTAACTTTAAAGAGGACATATATCATGACACCAAAATTTAAAAATTTTAAACATGACAAATTCGGGT  
CAAGTAGAAACGTATTTATTCGAAGTTATTTCCAAATACACCATGCTCGAAATAAAAAACAAAGTTTGGAAAGCTCTGCATAGCTAAGGTTAAGTTAGCATCGGAATCAATTGAACGCGCTTTGAA  
CCTTTGTAACGACAAATGTGATAGTTAGCGCCAAATGTCACGAAGAGGTATCTTCGCAACGATCTTGGCCATAAAGGCATTTAGTTCTGGAAGGAATCTGCCAAGCAATAAAGGGAGAGA  
TACTGTTAAGGTTAGCTGGTGTGAAGCAGATAAAGGATGCGATAAAGATCATTGGAGCTAGGAAAGGTGAAACCTAATAATATACCTTTGGAGTCTCCGACCTTCGAGTAGTCTTAAAGATTTT  
ATTTCAAGTTTAGTCTGAGGGAAGAGCCTCTCAAGCCTTCCGATGAAGATGAGGTTAAGAGAGGTTTCAAGAGGCGAGCAATAATCGAGGCCCTTAAGTCAAGCTTAATAATTTTGTTTAACT  
TTAAGAGGACATATATATGTCGAAGGAAGATAGTCATTGAGTTCCCTCGGAAGATATAGCAGAGGTGGTTTACACATCCGTTCTGTACGAGCAGCTTAGCGTCCCTTACAGGAGGAGTAGGG  
TGAATTTTAGGAGGGAAGGTAGGAGATAAGTCTGGAAGTTAGGCGCAACGATTCATCCGCGATGAGGGGACCGTTAACTCTTATCTCAGATGGATCAAGGTGGCCCTAGATGTTCTGAACATT  
CACCATCAGCATCACCATTAGGCTCGAGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATACTAGCATACCCCTTGGGGCTCTAAACGGGTCTTAGGGG  
TTTTTGGT

### Séquences :

#### T7 PROMOTEUR - LACO

TAATACGACTCACTATAGGGAATTGTGAGCGGATAACAATTCC

#### RIBOSOME BINDINGSITE

TTAAGCTTAAAGAGGAG

#### T7 TERMINATEUR

CTAGCATAACCCCTTGGGGCTCTAAACGGGTCTTAGGGGTTTTT

#### ETIQUETTE His-6:

5'-CATCCATCATCCATCAC-3' (HHHHHH)

#### ETIQUETTE STREPTAGII:

5'-TGG AGC CAC CCG CAG TTC GAA AAA-3' (WSHPQFEK)

SITES DE RESTRICTION UTILISES POUR LES CLONAGES :	
<i>Bgl</i> II	A/GATCT
<i>Bam</i> HI	G/GATCC
<i>Xba</i> I	T/CTAGA
<i>Sac</i> I	GAGCT/C
<i>Bsp</i> HI	T/CATGA (compatible <i>Nco</i> I)
<i>Hind</i> III	A/AGCTT
<i>Nco</i> I	C/CATGG
<i>Sal</i> I	G/TCGAC (compatible <i>Xho</i> I)
<i>Nde</i> I	CA/TATG
<i>Xho</i> I	C/TCGAG
<i>Blp</i> I(Bpu1102I)	GC/TNAGC
<i>Not</i> I	GC/GGCCGC
<i>Bsa</i> I(Eco31I)	GGTCTCN▼ CCAGAGNNNNN▲

## OLIGONUCLEOTIDES UTILISES POUR LE CLONAGE DES GENES PAR PCR

Protéine codées	Sites de restriction	Vecteur	Amorces	Matrice
<b>PaSua5<sup>-H</sup></b> <b>PaSua5<sup>-S</sup></b>	<i>Nde</i> I/ <i>Not</i> I	pET26b	Sua5F 5'-GGAATTCCATATGACTATCATAATAAACGTGAGGGAG Sua5 (HIS)R 5'- CCCCCCCCCGCGCCGCCTTAGTGATGGTGATGGTGATGAACCTTGATTATCCTA TACCCGAGGGCC Sua5 (STREP)R 5'- CCCCCCCCCTCGAGTTATTTTCGAACTGCGGGTGGCTCCAACTTTGATTAT CCTATACCCGAGGGCC	ADN génomique <i>Pyrococcus abyssi</i>
<b>PaCgi121<sup>-H</sup></b>	<i>Nde</i> I/ <i>Not</i> I	pET26b	Synthèse de gènes	
<b>PaPcc1<sup>-H</sup></b>	<i>Nde</i> I/ <i>Xho</i> I	pET26b	Sous clonage <i>Nde</i> I/ <i>Xho</i> I	pKEOPS 100
<b>PaBud32<sup>-H</sup></b>	<i>Nde</i> I/ <i>Not</i> I	pET9aSn1	Bud32F 5'-CATATG AAGCTAATCAAGCAAGG Bud32 (His)R 5'-GCGGCCGC TTA GTGATGGTGATGGTGATG CTGCGTAATCCAGCTTCTC	ADN génomique <i>Pyrococcus abyssi</i>
<b>PaKae1<sup>-H</sup></b>	( <i>Nco</i> I)/ <i>Not</i> I	pET28a	Cf Hecker et al., 2007	
<b>ScQri7<sup>-H</sup></b>	??	pET28a	Qri7F 5'-CCCCGAAGACACCATGAAAGTTTTAGCCATTGAGACTTCTTGC Qri7 (His)R 5'- CCCCCGCGCGCCGCTTAGTGATGGTGATGGTGATGTAGCTGGTCCGTTCTCCAA CC	ADN génomique <i>Saccharomyces cerevisiae</i>
<b>EcYgjD<sup>-H</sup></b>	<i>Bsa</i> I( <i>Nco</i> I)/ <i>No</i> <i>tl</i>	pET28a	YgjDF 5'-GGTCTCTCATGCGTGTACTGGGTATTGAACTTCC YgjD (His)R 5'- GC^GGCCGCTTAGTGATGGTGATGGTGATGCGCAGCCGGTAACCTCCGCCAGCGG CCAGCG	ADN génomique <i>Escherichia coli</i>
<b>EcYeaZ<sup>-H</sup></b>	<i>Nde</i> I/ <i>Not</i> I	pET9a	YeaZF 5'-GGAATTCCATATGCGAATTCTGGCTATCGATACCACAG YeaZ (His)R 5'-CCCCCCCCCGCGGCCGCTTAGTGATGGTGATGGTGATG TTCTTTGCCCGGAAGTTTCTTCATG	ADN génomique <i>Escherichia coli</i>
<b>EcYjeE<sup>-S</sup></b>	<i>Nde</i> I/ <i>Not</i> I	pET26b	Synthèse de gène	ADN génomique <i>Escherichia coli</i>
<b>EcYrdC<sup>-H</sup></b>	<i>Nde</i> I/ <i>Not</i> I	pET26b	YrdC (His)F 5'-GGAATTCCATATGAATAATAACCTGCAAAGAGACGC YrdC (His)R 5'- CCCCCCCCCGCGCCGCTTAGTGATGGTGATGGTGATGCCCTGTGCGAAACAGT TCACCCGTC	ADN génomique <i>Escherichia coli</i>
<b>P<sup>-H</sup>K</b>		pET28a	Digestion <i>Xba</i> I/ <i>Sal</i> I + « blunt »	pKEOPS 500
<b>KB<sup>-H</sup></b>	<i>Nde</i> I/ <i>Xho</i> I	pET26b (HisTag du vecteur)	KaeF 5'-GCGCGCCATATGCTAGCCCTGGGCATTGA BudR 5'-CCGCGGCTCGAGCTGCGTAATCCAGCTTCTCT	pKEOPS 500
<b>BC<sup>-H</sup></b>	<i>Nde</i> I/ <i>Xho</i> I	pET26b (HisTag du vecteur)	BudF 5'-GGCCCGCATATGAAGCTAATCAAGCAAGGAGCGG CgiR 5'-CGCGCGCTCGAGGAGGGCCTCGATTATTGCTG	pKEOPS 500
<b>KBC<sup>-H</sup></b>	<i>Nde</i> I/ <i>Xho</i> I	pET26b (HisTag du vecteur)	KaeF 5'-GCGCGCCATATGCTAGCCCTGGGCATTGA CgiR 5'-CGCGCGCTCGAGGAGGGCCTCGATTATTGCTG	pKEOPS 500
<b>P<sup>-H</sup>KB</b>				pKEOPS 500

-H = étiquette hexahistidine  
-S = étiquette StrepTagII

## OLIGONUCLEOTIDES UTILISES COMME AMORCES DE PCR POUR LA MUTAGENESE DIRIGEE

Nom	Mutation	Amorces de PCR	ADN matrice
<b>EcYgjD</b> <sup>(L134Y)-H</sup>	c400t - t401a (L134Y)	F' 5'-TTTCCGTTTGTGCGCTG <b>TATGTTTCCGGCGGTCATAC</b> -3' R' 5'-GTATGACCGCCGGAAC <b>ATACAGCGCAACAAACGGAAA</b> -3'	pET28a-YgjD(His)
<b>P<sup>-H</sup>K</b> <sup>(D159A)</sup>	a476c (D159A)	Kae1D159A-F AATCGGAAACGCCATAG <b>CTGTCTTTGCAAGGGAAC</b> Kae1D159A-R GTTCCCTTGCAAAGACA <b>GCTATGCGGTTTCCGATT</b>	pET28a-P <sup>-H</sup> K
<b>P<sup>-H</sup>K</b> <sup>(H107A)</sup>	c319g - a320c (H107A)	Kae1H107A-F ACCGATAGTTGGGGTAAAC <b>GCCTGTATAGCCCATGTAGAG</b> Kae1H107A-R CTCTACATGGGCTATACAG <b>GGCTTTACCCCACTATCGGT</b>	pET28a-P <sup>-H</sup> K
<b>B</b> <sup>(D127R)</sup> <b>C<sup>-H</sup></b>	g379c - a380g (D127R)	Bud32D127R-F GCGGGGATAGTTCACGGG <b>CGCTTAACGACCTCAAATAT</b> Bud32D127R-R ATATTTGAGTTCGTTAAG <b>CGCCCGTGAAGTATCCCCGC</b>	pET28a-BC <sup>H</sup>
<b>P<sup>-H</sup>(I5G-I7G)K</b>	a31g_t32g_a37g_t38g	5'-GGAGATATACATATGCAAGGAAAG <b>GGAGTCGGT</b> GAGTTCCCTCGGAAGATATAGC 5'-GCTATATCTTCCGAGGGGAAC <b>TACCCGACTCCCTTCCCTGCATATGTATATCTCC</b>	
<b>P<sup>-H</sup>(V61G-L65G)K</b>	t200g_c211g_t212g	5'-GCGATGAGGGGCACCG <b>GTA</b> ACTCTTAT <b>GGC</b> AGATGGATCAAGGTG-3' 5'-CACCTTGATCCATCTGCCATAAGAGTTACCGGTGCCCTCATCGC-3'	

-H = étiquette hexahistidine

-S = étiquette StrepTagII

## OLIGONUCLEOTIDES UTILISES POUR LA CONSTRUCTION ET LE CLONAGE DES GENES D'ARNt

Oligonucleotide	Sequence <sup>#</sup>
Pa_Val_GUC_f	5' AATTGGGCCCGTGGTCTAGACTGGTTATGACGCCACCCT <b>GACA</b> AGGTGGAGGTCGGGGGTTCAAATCCCCGCGGGCCCA <b>CCAT</b> GCATGCA
Pa_Val_GUC_r	5' TGGTGGGCCCGCGGGGATT <b>TGA</b> ACCCCGGACCTCCACCTTGTCAGGGTGGCGTCATAACCAAGTCTAGACCACGGGCC
Pa_Lys_AAA_f	5' AATTGGGCCCGTAGCTCAGCCTGGTCAGAGCACCGGGCTTT <b>TA</b> ACCCGGTGGTCGCGGGTTCAAATCCCCGCGGGCCCG <b>CCAT</b> GCATGCA
Pa_Lys_AAA_r	5' TGGCGGGCCGGCGGGGATT <b>TGA</b> ACCCCGGACCA <b>CCCGGGTTAA</b> AGCCCGGTGCTCTGACCAGGCTGAGCTACCGGCC
Pa_Ile_AUC_f	5' AATTGGGCCCGTGGTCTCAGCCTGGTTAGAGCGCCCGCT <b>GATA</b> AGCGGGAGGTCGGGGTT <b>CGA</b> AGCCCCGCGGGCCCA <b>CCAT</b> GCATGCA
Pa_Ile_AUC_r	5' TGGTGGGCCCGCGGGGCT <b>TCGA</b> ACCCCGGACCTCCCGCTTATCAGGCGGGCGCTCTAACCAGGCTGAGCCACGGGCC
Pa_Asn_AAC_f	5' AATTGCGCGCGTAGCTTAGCTGGTGGAGCGCCGGACT <b>GT</b> TAATCCGGCGGTCCCGGTT <b>CGA</b> ATCCGGCGGCGCG <b>CCAT</b> GCATGCA
Pa_Asn_AAC_r	5' TGGCGCGCGCGCCCGGATT <b>CGA</b> ACCGGGGACCGCGGATT <b>TA</b> ACAGTCCGGCGCTCCACCAGCTAAGCTACGGCGGC
Pa_Met_ATG_f	5' AATTGGGCCCGTAGCTCAGCCTGGTCAGAGCGCGGGCT <b>CATA</b> ACCGCGTGGTCCGGGGTTCAAATCCCCGCGGGCCCA <b>CCAT</b> GCATGCA
Pa_Met_ATG_r	5' TGGTGGGCCCGCGGGGATT <b>TGA</b> ACCCCGGACCA <b>CGCGGTTAT</b> GAGCGCGCGCTCTGACCAGGCTGAGCTACGGGCC
Ec_Val_GUC_f	5' AATTGCGTCCGTAGCTCAGTTGGTTAGAGCACCACT <b>TGACAT</b> GGTGGGGTCCGGTGGTT <b>CGA</b> GTCCACTCGGACGCACCA <b>CCAT</b> GCATGCA
Ec_Val_GUC_r	5' TGGTGGTCCGAGTGGACTCGAACCACCGACCCCA <b>CCAT</b> GTCAAGTGGTGTCTTAACCAACTGAGCTACGGACGC
Ec_Lys_AAA_f	5' AATTGGGTTCGTTAGCTCAGTTGGTAGAGCAGTT <b>GACTTTT</b> TAATCAATTGGTCGAGGTT <b>CGA</b> ATCCCTGCACGACCA <b>CCAT</b> GCATGCA
Ec_Lys_AAA_r	5' TGGTGGGTTCGTCAGGATT <b>CGA</b> ACCTGCGACCAATTGATT <b>AAA</b> AGTCAACTGCTCTACCAACTGAGCTAACGACCC
Ec_Ile_AUC_f	5' AATT <b>AGGCTTGT</b> AGCTCAGTGGTTAGAGCGCACCCCT <b>GATA</b> AGGGTGAGGTCGGTGGTT <b>CA</b> AGTCCACTCAGGCCTACCA <b>CCAT</b> GCATGCA
Ec_Ile_AUC_r	5' TGGTAGCCTGAGTGGACT <b>TGA</b> ACCACCGACCTCACCTTATCAGGGTGCGCTCTAACCACCTGAGCTACAAGCCT

<sup>#</sup> tRNA gene sequences are indicated in bold letters and the anticodon is underlined. EcoRI/PstI adaptor sequences are indicated in plain letters.

## OLIGONUCLEOTIDES UTILISES COMME SONDE POUR LES RETARDS SUR GEL

**RNA-F** GCUAUCUAGGAAUCUCGUUUUAUCAGACUGGAAUUAAGCGCGAGCUCGAAUAAGAGCUACUGUGG  
**RNA-R** CCACAGUAGCUCUUUAUCGAGCUCGCGCUUGAAUUCAGUCUGAUAAACGAGAUUCCAUGAUAGC

**DNA-F** GCTATCATGGAATCTCGTTTATCAGACTGGAATTCAAGCGCGAGCTCGAATAAGAGCTACTGTGG  
**DNA-R** CCACAGTAGCTCTTATTCGAGCTCGCGCTTGAATTCCAGTCTGATAAACGAGATTCCATGATAGC



# Annexe 3 - Séquences et propriétés des protéines

## *Pyrococcus abyssi* (GE5)

**Sua5 - PAB1302** (...= séquence du domaine Nter YrdC ; \_\_= séquence du domaine C ter SUA5, en *gras* : hélice alpha 8)

MTIIINVRERIEEWKIRIAAGFIREGKLVAFPTETVYGLGANALDENAVKRIFEAKGRPADNPLIIHIAFSFEQLEVLAKEIPEEAEMLAKEFW  
PGPLTLVLKSEVVRVITGGLDTVAVRMPAHEIALKLIELSERPIAAPSANISGKPSPTSAHHVAEDFYKIECIIDGGETRIGVESTVIDL  
TEWPPVLLRPGGL**PLEEIEKVI**GEIRIHPAVYGKSVDTAKAPGMKYRHYAPSAEVIVVEGPRDKVRRKIEELIAKFKEEGKKVGVIGSGSYDA  
DEVFYLGDTVVEE IARNLFKALRHMDRTGVDVILAEGVEEKGLGLAVMNRRLRKASGYRI IKV

## **Kae1 – PAB1159**

MLALGIEGTAHTLIGIGIVSEDKVLANVFDTLTTEKGGIHPKEAAEHARLMKPLLRKALSEAGVSLDDIDVIAFSQGPGLPALRVVATAARA  
LAVKYRKPIVGNHCIAHVEITKMFGVKDPVGLYVSGGNTQVLALEGGRYRVFGETLDIGIGNAIDVFARELGLFGPGPKVEKLAEKGEKYI  
ELPYAVKMDLSFSGLLTEAIRKYRSGKYRVEDLAYSFQETAFALVEVTERAVAHTKDEVVLVGGVAANNRLREMLRIMTEDRGIKFFVPP  
YDLCRDNGAMIAYTGLRMYKAGISFRLEETIVKQKFRTEVEIVW

## **Bud32 – PAB1047**

MKLIKQGAEAKIYLAEFSELYFDYPIKVIVKERIKKRYRIPEIDLKLRKERTIREARILRRAKEFGVNVVPYVFEVDTKNMIIVMEYIEGERLK  
ELLEKLPMEEERLKVCREVGRQIGKLHEAGIVHGDLTTSNMILREGKVYFIDFLAEFDDTIEAQGVDLHLLKRAMESTHYKWFERGFEEVLKG  
YIEIRGEDKGREIREKIREIELRGYRERSWITQ

## **Cgi121 – PAB1522**

MTPKFKNFKHDNSGQVETYLFEVISKYTMLEIKTKFGKLCIAKVKLASESIERALNLCNDNCQIVSAKCHEEVIFATILAIAKAFSSGRNTAKT  
IKGEILLRLAGVKIKDAIKIIGARKGENLIIFYGVSDPCSSLKDFISSLGLREEPLKPCDEDEVKRGFERAAIIEAL

## **Pcc1 – PAB3073**

MQGKIVIEFPSEDIAEVVYTSVLVEYHVSVPYRRSRVNFRRREGRRIVLEIEANDSSAMRGTVNSYLRWIKVALDVINI

## *Saccharomyces cerevisiae* (S288c)

**ScQri7** (*souligné : signal d'adressage mitochondrial. Pointillé : 30 premiers aa, Pointillé + plein : 33 aa*)

MSIQKCTGRFLLDNYRIWORPAFNRPIOLRKGYKVLAIETSCDDTCVSVLDRFSKSAAPNVLANLKDTLDSIDEGGIPTKAHIHHQARIGPL  
TERALIESNAREGIDLCVTRGPGMPGSLSGGLDFAKGLAVAWNKLPIGVHHMLGHLLIPRMGTNGKVPQFPFVSVLLVSGGHTTFVLSRAIDD  
HEILCDTIDIAVGDSLDCGRELFGKGTMIAREMEKFINQDINDQDFALKLEMPSPLNKSASKRNLSFSFSAFITALRTNLTKLGKTEIQEL  
PEREIRSIAYQVQESVFDHIINKLKHVLKSQPEKFNKVRREFVCSGGVSSNQRLRTKLETGLTNSTSFNFYPPMDLCSDNSIMIGWAGIE  
IWESLRLVSDLDICPIRQWPLNDLLSVDGWRDQL

## **Sua5**

MYLGRHFLAMTSKALFDTKILKVNPLSIIFSPDAHIDGSLPTITDPETEALVEAARIIRDDETVAFPPTETVYGLGGSALNDNSVLSIYRAK  
NRPSDNPLITHVSSIDQLNRKVFNQPHLSGTSLFDNIPSIYRPLISSLWPGPLTIILLVPVSSEHSALSKLTTADQPTFAVRI PANPVARALIA  
LSDTPIAAPSANASTRPSPTLASHVYHDLKDKIPIILDGGACKVGVESTVVDGLCNPTLLRPGGFTYEEIVKLGGEAWSLCKVENKKTVEKG  
EKVRTPGMKYRHYSPSAKVLLVPHCEGDGILKGVDRMERLKLRIETELKANSNIKKIALTSLKLKRLSDLSQSKIFNEPDFSSKTFIIERLGQ  
SGEEIQTNLFAALRKVDENDKVDLIFVEGINEEGEGLAVMNRRLRKAANNCIQF

## **Kae1**

MVNLNTIPKNGRDYIYALGLEGSANKLVGVIVKHPLLPKHANSLSYDCEAEMLSNIRDYVTPPGEGFLPRDTRHHRNWCIRLIKQALAE  
ADIKSPTLDDIDVICFTKGPGMGAPLHVVIAARTCSLLWDVPLVGNHCIGHIEMGREITKAQNPVVLYVSGGNTQVIAYSEKRYRIFGETLD  
IAIGNCLDRFARTLKI PNEPSPGYNIEQLAKKAPHKENVELPYTVKGMDSLMSGILASIDLLAKDLFGKNKKKILFDKTTGEQKVTVEDLC  
YSLQENLFAMLVEITERAMAHVNSNQVLIVGGVGCNVRQLQEMMAQMCKDRANGQVHATDNRFCDINGVMIAQAGGLEMYRMGGIVKDFSETVVT  
QKFRTEVYAAWRD

## **Bud32**

MTQEFIDKVSSYLTPDVDIAPISQGAEAIVFTTTTHPYLPRAKDSHQYI IKYRPPKRYRHPQIDQALTKHRTLNESRLLAKLYLIPGLCVPO  
LIACDPYNGFIWLEFLGEDLPGGHGFSNLKNFLWMHDQDPYSDLVATTLRKVGRQIGLLHWN DYCHGDLTSSNIVLVRD GARWTPH LIDFGLG  
SVSNLVEDKGV DLYVLERAILSTHSHKAEKYN AWIMEGFEEVYREQGAKGAKKLKEVTKRFEFVRLRGRKRSMLG

## **Cgi121**

MVVSII PQFPDIKVS LALFEQVKNKEIRSKMSELSTSF AFIDPRLVCSGEQMYSAIYKTLIEVKYNKMRTRNLNSECVLCLSP TSNISDAFL  
KFGIKDDSSQLICLKFHTNTDDVDKEQLRTIMTSIVKGQIEFNDNLSRFYDEALIRKIYKLSDDFKPQDVNGLSRALVD AIQLRGV

## **Pcc1**

MTSKREKSLDHTLELKI PFETERQATIATKVLS PDPILKPQDFQVDYSSEKNVMLVQFRSIDDRVLRVG VSSIIDS IKTIVEAMDVLS

## **Gon7**

MKL PVAQYSAPD GVEKSFAPIRDDPRYMTTEGR TTGPS DHVLNAGQIDRDKPSEP ERTKDGSQ LTYLGQLRTQLTGLQDDINEFLTGRMELAK  
NKKKAGADEKRIQEEINQLDGGDGEDAV

## Escherichia coli (K12 -MG1665)

### YrdC

MNNNLQRDAIAAAIDVLNEERVIAYPTEAVFVGCDPDESETAVMRLELQKRPVDKGLILIAANYEQLKPYIDDTMLTDVQRETIFSRWPGPV  
TFVFPAPATTPRWLTGRFDSLAVRVTDHPLVVALCQAYGKPLVSTSANLSGLPPCRTVDEVRAQFGAAFPVVPGETGGRLNPSEIRDALTGEL  
FRQG

### YgjD

MRVLGIETSCDETGIAIYDDEKGLLANQLYSQVKLHADYGGVVPPELASRDHVRKTVPLIQAAALKESGLTAKDIDAVAYTAGPGLVGLLVGAT  
VGRSLAFWDVPAIPVHHMEGHLLAPMLEDNPPPEFFVALLVSGGHTQLISVTGIGQYELLGESIDDAAGEAFDKTAKLLGLDYPGGPILLSKM  
AAQGTAGRFVFPMPMDRPLDLSFSGSLKTFAANTIRDNGTDDQTRADIARAFEDAVVDTLMIKCKRALDQTGFKRLVMAGGVSANRTLRAKL  
AEMMKRRRGEVFFYARPEFCTDNGAMIAYAGMVRFKAGATADLGVSVRPRWPLAELPAA

### YeaZ

MRILAITATEACSVALWNDGTVNAHFELCPREHTQRILPMVQDILTTSGTSLTDINALAYGRGPGSFTGVRIGIGIAQGLALGAELPMIGVS  
TLMTMAQGAWRKNAGATRVLAIDARMGEVYWAQYQDENGIVHGEETEAVLKPEIVHERMQQLSGEWVTVGTGWQAWPDLGKESGLVLRDGEV  
LLPAAEDMLPIACQMFAEGKTVAVEHAEPVYLNNVAWKKLPGE

### YjeE

MMNRVILPLDEQATLDLGERVAKACDGATVIYLYGDLGAGKTTFSRGFLQALGHQGNVKSPTYTLVEPYTLNLMVYHFDLYRLADPEELEFM  
GIRDYFANDAICLVEWPPQGTGVLDPDVEIHIDYQAQGREARVSAVSSAGELLARLAG

## TABLE DES PROPRIETES BIOCHIMIQUES DES PROTEINES

	Amino acid	Molecular mass (Da)	1g/L = $\mu$ M	pI	Ext coeff	Abs (1g/L)
<b><i>Escherichia coli (K12 MG1655)</i></b>						
YrdC	190	20.768	48 $\mu$ M	4,94	16960	0,817
YeaZ	231	25.181	39,7 $\mu$ M	4,99	49960	1,984
YgjD	337	36.008	27,8 $\mu$ M	5,92	22920	0,637
YjeE	153	16.853	59,3 $\mu$ M	4,49	17420	1,034
<b><i>Pyrococcus abyssi (GE5)</i></b>						
Sua5(His)	346	38.268	26,1 $\mu$ M	6,38	28420	0,743
Sua5 $\Delta$ Nter-Strep	119	13.535	73,9 $\mu$ M	9,12	12950	0,957
Sua5 $\Delta$ Cter-Strep	217	23.918	41,8 $\mu$ M	5,26	24980	1,044
Sua5 $\Delta$ Cter $\Delta$ H8-Strep	206	22.696	44,1 $\mu$ M	5,47	24980	1,101
Pcc1-(His)	83	9.779	102,3 $\mu$ M	8,22	11460	1,172
Kae1	324	35.368	28,3 $\mu$ M	6,68	21890	0,619
Bud32	220	26.343	38 $\mu$ M	8,95	25900	0,938
Cgi121	171	19.030	52,6 $\mu$ M	8,85	4470	0,235
KEOPS-(His)	798	90.466	11 $\mu$ M	8,74	63720	0,704
<b>Sous-complexes de <i>P.abyssi</i></b>						
Pcc1(His)KAE1	407	45.147	22,1 $\mu$ M	7,26	33950	0,741
Bud32-Cgi121-(His)	397	~46.000	21,7 $\mu$ M	8,97	30370	0,658
Kae1- Bud32-(His)	550	~62.400	16 $\mu$ M	8,39	47790	0,764
Kae1-Bud32-Cgi121-(His)	721	~81500	12,2 $\mu$ M	8,69	52260	0,641
<b><i>Saccaromyces cerevisiae</i></b>						
Sua5	426	46.530	21,5 $\mu$ M	6,77		
Kae1	386	42.740	23,4 $\mu$ M	6,79		
Bud32	261	29.930	33,4 $\mu$ M	8,81		
Cgi121	181	20.660	48,4 $\mu$ M	5,91		
Gon7	123	13.600	73,5 $\mu$ M	4,72		
Pcc1	88	9.990	100 $\mu$ M	5,3		
KEOPS	1039	116.900	8,6 $\mu$ M	6,31		
Qri7-(His)	381	42.446	23,2 $\mu$ M	6.24	32470	0.753

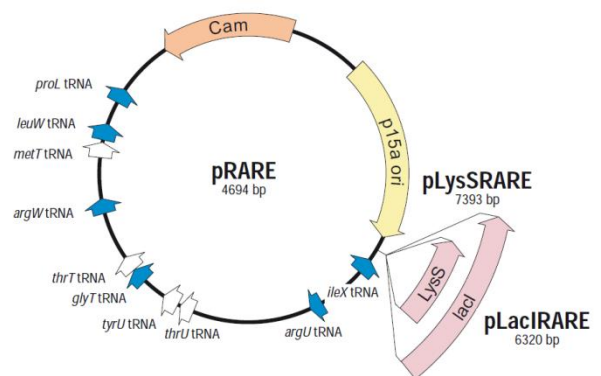
## DESCRIPTION DE LA SOUCHE ROSETTA2(DE3)PLYSS (NOVAGEN)

Génotype:  $F^- ompT hsdS_B(r_B^- m_B^-) gal dcm$  (DE3) pLysSRARE2 (CamR)

La souche **Rosetta2** est un dérivé de la souche BL21 (déficiente des protéases Lon et OmpT), conçue pour améliorer l'expression des protéines eucaryotes qui contiennent des codons rarement utilisés dans *Escherichia coli*. La souche originelle Rosetta fournit les ARNt pour les codons AUA (ile), AGG (arg), AGA (arg), CUA (leu), CCC (pro) et GGA (gly) sur un plasmide compatible résistant au chloramphénicol, pRARE. La souche Rosetta 2 fournit un ARNt pour un 7<sup>ème</sup> codon rare CGG (arg) en plus des six précédents. En fournissant ces codons rares, les souches Rosetta permettent une traduction « universelle »

Le terme **DE3** indique que l'hôte est un lysogène du phage  $\lambda$ DE3, et porte ainsi une copie chromosomique du gène codant la T7 ARN polymérase sous contrôle du promoteur lacUV5 (inducible by addition of IPTG or using an auto-inducible medium). Cette souche est utilisable pour la production de protéines dont les gènes sont clonés dans des vecteurs pET et inducibles par IPTG.

Dans le dérivé **pLysS**, les gènes d'ARNt (*argU*, *argW*, *ileW*, *glyT*, *leuW*, *proL*, *metT*, *thrT*, *tyrU*, *thrU* et *argX*) sous contrôle de leur promoteur d'origine sont présents sur le même plasmide qui porte le gène codant le lysozyme du phage T7 : le plasmide **pLysRARE2**. Le lysozyme du phage T7 est un inhibiteur naturel de la T7 ARN polymérase, et supprime ainsi l'expression basale de cette protéine avant l'induction, stabilisant ainsi les plasmides pET recombinants codant des protéines qui affectent la croissance et la viabilité cellulaire.



Ce plasmide confère une résistance au chloramphénicol (**CamR**)

### Autres souches utilisées:

#### BL21(DE3)

Génotype:  $F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) \lambda$ (DE3)

#### XL1 BLUE (Stratagène)

Génotype:  $endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F' [::Tn10 proAB^+ lacI^q \Delta(lacZ)M15] hsdR17(r_K^- m_K^+)$

#### TOP10 (Invitrogen)

Génotype:  $F^- mcrA \Delta(mrr-hsdRMS-mcrBC) \phi80lacZ\Delta M15 \Delta lacX74 nupG recA1 araD139 \Delta(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 \lambda^-$

Site complet pour comprendre les sigles de génotypes : [http://openwetware.org/wiki/E.\\_coli\\_genotypes](http://openwetware.org/wiki/E._coli_genotypes)

# ANNEXE 4 : Recettes de tampons

## RETARD SUR GEL (NATIF)

<u>TAMPON DE MIGRATION TEG 5X</u>			
<i>Pour 1L</i>			
Tris	30,3g	(250mM)	
Glycine	3g	(40mM)	
EDTA 0,5M	1mL	(0,5mM)	
H2O	qsp	1L	
Filtrer			
Pas d'ajustement du pH (environ 9/10)			

<u>TAMPON DE DEPOT</u>			
<i>Pour 1mL</i>			
Tris 1MpH 7.5	10µL	(10mM)	
Glycérol 50%	400µL	(20%)	
EDTA 0,5M	2µL	(1mM)	
BSA50mg/mL	2µL	(100µg/mL)	
Xylène cyanol	1mg	(1mg/mL)	
H2O	qsp	1mL	

## SDS-PAGE (DENATURANT)

<b>GEL DE POLYACRYLAMIDE POUR PROTEINES</b>							
Gel de Running (Lower), 20mL final (4 petits gels)							
	H <sub>2</sub> O	Tris 1,5M pH8,8	Glycérol 50%	SDS 10%	Acryl/Bis 37/1 40%	APS 10%	Temed
5%	9,5mL	5 mL	2,5 mL	200 µl	2,7mL	200µL	20 µL
7%	8,4mL	5 mL	2,5 mL	200 µl	3,8mL	200µL	20 µL
7,5%	8,1mL	5 mL	2,5 mL	200 µl	4,1mL	200µL	20 µL
8%	8mL	5 mL	2,5 mL	200 µl	4,3mL	200µL	20 µL
9%	7,3mL	5 mL	2,5 mL	200 µl	4,9mL	200µL	20 µL
10%	6,8mL	5 mL	2,5 mL	200 µl	5,4mL	200µL	20 µL
12%	5,7mL	5 mL	2,5 mL	200 µl	6,5mL	200µL	20 µL
15%	4,7mL	5 mL	2,5 mL	200 µl	8,1mL	200µL	20 µL
18%	2,6mL	5 mL	2,5 mL	200 µl	9,7mL	200µL	20 µL
Gel de Stacking (Upper), 10mL final (4 petits gels)							
	H <sub>2</sub> O	Tris 1,5M pH8,8	Glycérol 50%	SDS 10%	Acryl/Bis 37/1 40%	APS 10%	Temed
3%	5,3	1,25	2,5	100µL	810µL	100µL	10µL

<u>TAMPON DE MIGRATION 10X :</u>			
<i>Pour 1L</i>			
Tris	30 g		
Glycine	144 g		
SDS 20%	50mL		
H <sub>2</sub> O	qsp	1L	

<u>LAEMMLI 5X (TAMPON DE DEPOT)</u>			
<i>Pour 16mL</i>			
H <sub>2</sub> O	6,8mL		
TrisHCl pH 6.8 0,5M	2mL		
Glycérol 100%	3,2mL		
SDS 20%	1,6mL		
bmercaptoethanol	0,8mL		
Bleu de Bromophenol 1%	1,6mL		

## SEPARATION ET EXTRACTION DES ARN SUR GEL

### **Tampon de dépôt 2X Formamide (pour gel acrylamide urée)**

95% Formamide  
18mM EDTA  
0,025% SDS  
0,025% Bromophénol  
0,025% Xylène Cyanol

### **Tampon Crush and Soak (pour extraction d'ARNt sur gel)**

0,5M acétate d'ammonium  
10mM acétate de magnésium  
1mM EDTA  
0,1% SDS

## **CHROMATOGRAPHIE SUR COUCHE MINCE**

### **POUR SEPARATION AMP, ADP, ATP, PI :**

LiCl 0,5M, formic acid 1M  
KH<sub>2</sub>PO<sub>4</sub> 0,5M, pH3,5 (buffered with orthophosphoric acid)

### **POUR SEPARATION PNUCLEOTIDES EN 2 DIMENSION :**

#### **A/N1= isobutyric acid/concentrated ammonia/eau (66/1/33 [v/v/v])**

165mL d'acide isobutyric  
2,5mL de solution ammonia (25% concentré)  
82,5mL d'H<sub>2</sub>O.

#### **B/R2= phosphate buffer / NH<sub>4</sub> sulfate / n-propanol (100/60/2 [v/w/v]).**

150g de ultrapure NH<sub>4</sub> sulfate  
250mL de 100mM de Na phosphate buffer (pH 6.8)  
Mélanger bien et chauffer à 50/60°C jusqu'à ce que le NH<sub>4</sub> soit entièrement dissous. Laisser la solution refroidir à T° ambiante,  
+5mL de 1-propanol (ce n'est pas de l'isoprop !!!).  
Continuer à mélanger jusqu'à ce que le précipité disparaisse.

#### **C/N2 = isopropanol/HCl /eau (68/18/14 [v/v/v]).**

170mL d'isopropanol (2propanol)  
45mL de HCl fumant (37% le plus pure possible)  
35mL d'H<sub>2</sub>O.

pH table for phosphate buffers<sup>a</sup>:

Desired pH <sup>a</sup>	sodium or potassium, phosphate mono salt solution (ml)	sodium or potassium, phosphate di salt solution (ml)	H <sub>2</sub> O (ml)	Desired pH <sup>a</sup>	sodium or potassium, phosphate mono salt solution (ml)	sodium or potassium, phosphate di salt solution (ml)	H <sub>2</sub> O (ml)
5.7	93.5	6.5	100	6.9	45.0	55.0	100
5.8	92.0	8.0	100	7.0	39.0	61.0	100
5.9	90.0	10.0	100	7.1	33.0	67.0	100
6.0	87.7	12.3	100	7.2	26.0	72.0	100
6.1	85.0	15.0	100	7.3	23.0	77.0	100
6.2	81.5	18.5	100	7.4	19.0	81.0	100
6.3	77.5	22.5	100	7.5	16.0	84.0	100
6.4	73.5	26.5	100	7.6	13.0	87.0	100
6.5	68.5	31.5	100	7.7	10.5	90.5	100
6.6	62.5	37.5	100	7.8	6.5	93.5	100
6.7	56.5	43.5	100	7.9	7.0	93.0	100
6.8	51.0	49.0	100	8.0	5.3	94.7	100

<sup>a</sup> It is strongly recommended to check the final pH with a pH meter.

**TOUS LES SOLVANTS SONT GARDES A TEMPERATURE AMBIANTE DANS UNE ZONE VENTILEE.**

Le solvant A pue et le C corrode.



# Summary

Transfer RNA are central elements of the translational system and carry a large diversity of modified nucleosides (derived from canonical nucleosides A, U, G, and C), which tune the stability, the decoding capacity and the identity of these oligonucleotides.  $t^6A$  (threonylcarbamoyl- $N^6$ -adenosine) is a hypermodified nucleoside found at the position 37 (next to the anticodon) in all tRNA decoding ANN codons. It plays an essential role in the fidelity of translation through two main functions: (i) it ensures a correct conformation of the anticodon loop; (ii) it enhances codon/anticodon pairing to prevent frameshifting during translation. This nucleoside is universal, found in Archaea, Bacteria, Eukarya and also in organites such as mitochondria, which suggests that it appeared early in the evolution, probably before the last universal common ancestor (LUCA). Despite the importance of  $t^6A$  and its distribution, its biosynthetic pathway has remained unknown for almost 40 years.

Recently, genetic studies have shown that two universal proteins, Sua5/YrdC and Kae1/YgjD, are both necessary for synthesis of  $t^6A$  in *Saccharomyces cerevisiae* and *Escherichia coli*. In Bacteria, the *in vitro* synthesis of  $t^6A$  requires two other bacterial specific proteins called YeaZ and YjeE. In Archaea and Eukarya, Kae1 (the YgjD orthologue) is a part of a conserved protein complex called KEOPS (for Kinase Endopeptidase and Other Proteins of Small size), with three other proteins Bud32, Cgi121 and Pcc1, that have no bacterial homologues. Since its discovery in 2006 in yeast, this complex has been involved in several cellular processes (telomere homeostasis, genome maintenance, transcription regulation), but its real function remained unclear.

Using an *in vitro* biochemical approach we aimed to characterize and compare the  $t^6A$  biosynthesis systems from the three domains of life, using as model organisms *Pyrococcus abyssi* (Archaea) *Saccharomyces cerevisiae* (Eukarya), and *Escherichia coli* (Bacteria). We have reconstituted for the first time an *in vitro* system for  $t^6A$  modification in Archaea and Eukarya, using purified KEOPS and Sua5. This allowed us to propose a model for the catalytic mechanism, and using *in vitro* complementation experiments we demonstrated that this mechanism is universal: Sua5/YrdC orthologues are interchangeable, and the KEOPS complex is the functional analogue of the bacterial trio YeaZ/YgjD/YjeE. In the second part of this work we have studied the role of each sub unit in the synthesis of  $t^6A$ . Using KEOPS from *P. abyssi* as model we demonstrated that Kae1 is the only catalytic component while the three other partners have distinct functions in dimerization, tRNA binding and allosteric regulation. Finally, we have focused on the  $t^6A$  synthesis in the mitochondria of *S. cerevisiae*, and shown that Sua5 and Qri7, the mitochondrial orthologue of Kae1/YgjD, catalyze together the synthesis of  $t^6A$  and so represent a minimal two-component system.

Overall these findings shed light on the reaction mechanism of  $t^6A$  synthesis in the three domains of life, and allowed proposing a scenario concerning the history of the  $t^6A$  synthesis machinery and its evolution.

**Key words:**  $t^6A$ , tRNA, LUCA, Archaea, universal protein, Kae1, Sua5, KEOPS.

# Résumé

Les ARN de transfert, éléments centraux de la traduction, présentent une grande variété de nucléosides modifiés dérivés des nucléosides canoniques (A, U, G et C), qui modulent la stabilité, la capacité de décodage et l'identité de ces molécules.  $t^6A$  (thréonylcarbamoyl- $N^6$ -Adénosine) est un nucléoside hypermodifié retrouvé en position 37 (adjacent à l'anticodon) au niveau de tous les ARNt qui s'apparient aux codons de la forme ANN. Il joue un rôle essentiel dans la fidélité de traduction à travers deux fonctions principales : (i) il intervient dans le maintien de la bonne conformation de la boucle anticodon ; (ii) il facilite l'appariement codon/anticodon afin d'éviter le décalage de cadre de lecture durant la synthèse protéique. Ce nucléoside modifié est universel, présent chez les Archées, les Bactéries, les Eucaryotes, mais également chez les organites (mitochondries et chloroplastes), ce qui suggère que son apparition représente une acquisition évolutive importante et très ancienne, probablement antérieure au dernier ancêtre commun universel (LUCA). Pourtant, la voie de biosynthèse de  $t^6A$  est restée inconnue pendant près de quarante ans.

Récemment, des études de génétique ont montré que deux protéines universelles, Sua5/YrdC et Kae1/YgjD, sont nécessaires à sa synthèse chez *Saccharomyces cerevisiae* et *Escherichia coli*. Chez les Bactéries, la synthèse *in vitro* de  $t^6A$  requiert la présence de deux autres protéines spécifiques à ce domaine du vivant : YeaZ et YjeE. Chez les Archées et les Eucaryotes, Kae1 (l'orthologue de YgjD) fait partie d'un complexe protéique conservé appelé KEOPS (pour Kinase Endopeptidase and Other Proteins of Small size), aux côtés de trois autres protéines : Bud32, Cgi121 et Pcc1, qui n'ont pas d'homologues chez les Bactéries. Depuis sa découverte en 2006 chez *S.cerevisiae*, ce complexe a été impliqué dans plusieurs processus cellulaires (homéostasie des télomères, maintien du génome, régulation de la transcription), sans que sa fonction ne soit clairement élucidée.

Nous avons entrepris de caractériser et de comparer par une approche biochimique *in vitro* les machineries de biosynthèse de  $t^6A$  issues des trois domaines du vivant, en utilisant comme organismes modèles l'Archée *Pyrococcus abyssi*, l'Eucaryote *Saccharomyces cerevisiae* et la Bactérie *Escherichia coli*. (i) Nous avons montré pour la première fois que le complexe KEOPS et la protéine Sua5 catalysent ensemble la synthèse de  $t^6A$  chez les Archées et les Eucaryotes. Nos résultats nous ont permis d'élaborer un modèle de mécanisme catalytique, et nous avons montré par des expériences de complémentation *in vitro* que ce mécanisme est universel : les différents orthologues Sua5/YrdC sont interchangeable, et le complexe KEOPS est l'analogue fonctionnel du trio de protéines YgjD/YeaZ/YjeE Bactérien. (ii) Nous avons alors étudié le rôle de chacune des sous-unités du complexe KEOPS de *Pyrococcus abyssi* dans la synthèse de  $t^6A$ . Ainsi, nous avons montré que Kae1 est le seul composant catalytique *stricto sensu* et que les trois autres partenaires ont des fonctions distinctes dans la régulation de l'activité catalytique. (iii) Enfin, nous avons étudié la synthèse de  $t^6A$  chez la mitochondrie de *S.cerevisiae*, et avons montré que Sua5 et la protéine Qri7, l'orthologue mitochondrial de Kae1/YgjD, catalysent ensemble la synthèse de  $t^6A$  et constituent ainsi un système minimaliste à deux composants.

Ces résultats ouvrent la voie à une compréhension détaillée du mécanisme de biosynthèse de  $t^6A$  dans les trois domaines du vivant, et permettent de proposer des scénarii évolutifs concernant l'histoire de la machinerie de synthèse de ce nucléoside modifié universel.

**Mots clés :**  $t^6A$ , ARNt, LUCA, Archées, protéine universelle, Kae1, Sua5, KEOPS.